BAYESIAN CAUSAL INFERENCEE IN META-ANALYSIS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

JINCHENG ZHOU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISED BY DR. HAITAO CHU

May, 2019

Acknowledgements

First and foremost, I would like to express my eternal gratitude to my advisor Dr. Haitao Chu for his tremendous mentorship, support and guidance. I feel extremely fortunate to have connected with Dr. Chu even before I entered the M.S. program, and to have worked with him since I became a master student. He has served as an excellent role model, provided a never-ending supply of patience, and has guided me all the way to finally become a qualified and confident PhD graduate. He's been always willing to share his own experience, either successful or unsuccessful, to provide hands-on guidance whenever I met challenges regarding research and beyond. I am also extremely appreciative of all the support and insights from Dr. James Hodges, who is a co-author of all my papers related to my thesis and is serving on my dissertation committee. He has provided many insightful suggestions on my research projects, as well as various help on improving my scientific writing and presentation skills. I especially appreciate the informative discussions he provided when we worked on the *Biometrics* paper revision. My sincere thanks also go to the rest of my thesis committee, Dr. Julian Wolfson and Dr. Richard Maclehose, for taking the time to review my proposal and thesis. Their helpful comments, questions and encouragement have made this dissertation stronger and more complete.

Additionally, I would like to extend my gratitude to Dr. Chu's all previous students for their help with many techniques and programming details. Thanks to my office-mates and classmates for making this challenging process of pursuing a PhD much more enjoyable: we shared ideas, helped each other to overcome difficulties, and established friendship during the time we spent together. In the meantime, I gratefully acknowledge

the Doctoral Dissertation Fellowship that financially supported my research in the past academic year.

Lastly, without the support of my family I could not have been where I am today. I am deeply and forever indebted to my parents for the utmost care, love and unwavering support they have given me. I would also like to express gratitude to my parents-in-law, and my grandparents for their encouragement and emotional support. I am grateful that my auntie Sharon Jia and uncle Sam Xiao motivated me to study biostatistics abroad. They have offered me tremendous help to make my life as a graduate student in the United States smoother. Special thanks to my husband Peng Xu, who has been an endless source of strength for me over the years. He has always been there to witness all my ups and downs of my whole PhD journey. His unconditional love and continuous emotional support are among the most important reasons for what I have accomplished.

Dedicated in honor of

My parents Mingdong Zhou, Min Jia, and my beloved husband Peng Xu

for their infinite love, trust and support.

Abstract

While the randomized clinical trial (RCT) is the gold standard for investigating the effect of a medical intervention, noncompliance to assigned treatments can threaten a trial's validity. Noncompliance, if not appropriately controlled, can introduce substantial bias into the estimate of treatment effect. The complier average causal effect (CACE) approach provides a useful tool for addressing noncompliance, where CACE measures the effect of an intervention in the latent subgroup of the study population that complies with its assigned treatment (the compliers). Meta-analysis of RCTs has become a widely-used statistical technique to combine and contrast results from multiple independent studies. However, no existing methods can effectively deal with heterogeneous noncompliance in meta-analysis of RCTs. For example, the commonly used meta-analysis regression methods investigate the impact of study-level variables (e.g., mean age of the study population) on the study-specific treatment effect size by assuming the study-level covariates to be fixed. However, noncompliance rates generally differ between treatment groups within a study and are commonly considered as random rather than fixed post-randomization variables. In addition, noncompliance may dynamically interact with the primary outcome and thus affect the response to treatment. Thus, meta-regression methods are not suitable to controlling for noncompliance.

This thesis focuses on developing Bayesian methods to estimate CACE in meta-analysis of RCTs with binary or ordinal outcomes. Bayesian hierarchical random effects models are developed to appropriately account for the inherent heterogeneity in treatment effect and noncompliance between studies and treatment groups. We first present a Bayesian hierarchical model to estimate the CACE where heterogeneous compliance rates are available for each study. Second, we extend our approach to deal with incomplete noncompliance when some RCTs do not report noncompliance data. The results are illustrated by a re-analysis of a meta-analysis comparing the effect of epidural analgesia in labor versus no or other analgesia in labor on the outcome cesarean section, where noncompliance varies substantially between studies. Simulations are performed to

evaluate the performance of the proposed approach and to illustrate the importance of including appropriate random effects by showing the impact of over- and under-fitting.

Furthermore, we develop an R package, **BayesCACE**, to provide user-friendly functions to implement CACE analysis for binary outcomes based on the proposed Bayesian hierarchical models. This package includes flexible functions for analyzing data from a single RCT and from a meta-analysis of multiple RCTs with either complete or incomplete noncompliance data. The package also provides various functions for generating forest, trace, posterior density, and auto-correlation plots, and to review noncompliance rates, visually assess the model, and obtain study-specific and overall CACEs.

Table of Contents

Acknowledgements i

Abstract iv

List of Tables ix

List of Figures xi

1 Introduction 1

1.1 Noncompliance and the causal effect in a single study 1

1.2 Causal effects in meta-analysis 2

1.3 Noncompliance with missing data in a single study 3

1.4 Handling missing data and noncompliance in meta-analysis 4

1.5 R packages to perform CACE analysis 5

2 Estimating CACE in Meta-analysis with Noncompliance 7

2.1 Statistical Methods 8

2.1.1 Notation and assumptions 8

2.1.2 Estimation and inference for the causal effect 10

2.1.3 Likelihood and the Bayesian hierarchical model 11

2.1.4 Accounting for heterogeneity through random effects models . . 12

2.1.5 Implementation 13

2.2 Case Study 13

2.2.1 Data and descriptive analysis . . . 13
2.2.2 Likelihood and priors . . . 15
2.2.3 Model selection and results . . . 17
2.3 Simulation Studies . . . 21
2.4 Discussion . . . 24
2.5 Supporting Information . . . 25

3 A Bayesian Hierarchical CACE Model Accounting for Incomplete Noncompliance Data in Meta-analysis 27
3.1 A Motivating Study . . . 28
3.1.1 Data Sources . . . 28
3.1.2 Analysis of Event Rates and Noncompliance Rates . . . 30
3.2 Statistical Methods . . . 33
3.2.1 Definition of the Complier Average Causal Effect (CACE) . . . 33
3.2.2 Estimation and Inference . . . 38
3.3 Case Study Results . . . 42
3.3.1 Model Selection Results . . . 42
3.3.2 Sensitivity to the LI Assumption . . . 49
3.4 Simulation . . . 52
3.4.1 Simulation Setups . . . 52
3.4.2 Simulation Results . . . 54
3.5 Discussion . . . 59

4 Estimating Causal Effect using the Bayesian Method with the R Package BayesCACE 61
4.1 CACE, assumptions and definition . . . 62
4.2 Estimating CACE . . . 64
4.2.1 CACE for a single trial with noncompliance . . . 64
4.2.2 CACE for a meta-analysis with complete compliance information 66
4.2.3 CACE for meta-analysis with incomplete compliance information 68
4.3 Using the R package **BayesCACE** . . . 70

4.3.1 Data structure for estimating the CACE 70
4.3.2 Plotting noncompliance rates 72
4.3.3 CACE analysis for a single study or in a meta-analysis 73
4.3.4 Plotting the trace plot, posterior density, and auto-correlation . . 86
4.3.5 Plotting the study-specific CACE in a forest plot 90
4.4 Discussion 92

5 Conclusion **95**
5.1 Summary of major findings 95
5.2 Future research 96

References **106**

Appendix A. Supplementary Materials for Chapter 2 **107**
A.1 Data display for the case study 107
A.2 Parameter estimates of the case study 108
A.3 Sample R code 109
A.4 Sensitivity analysis on priors in the case study 111
A.5 Additional simulation results 112

List of Tables

2.1 Observed groups, latent compliance classes and outcome probabilities of trial i . . . 9
2.2 Selection of random effects using a forward selection procedure on RCTs of epidural analgesia in labor . . . 18
2.3 Estimated probability of selecting each candidate model using DIC*, based on 2000 simulated datasets . . . 22
2.4 Performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets . . . 23
3.1 Data from randomized controlled trials of epidural analgesia in labor . . . 29
3.2 Observed groups, latent compliance classes and outcome probabilities of trial i . . . 36
3.3 Typical data for study i with missing actual treatment received status in randomization group $r \in \{0, 1\}$ . . . 37
3.4 Selecting random effects using a forward selection procedure for the epidural analgesia in labor meta-analysis . . . 44
3.5 Summary of parameter estimates for the epidural analgesia meta-analysis 46
3.6 Simulation results: relative bias (ReBias), mean square error (MSE), 95% CI coverage probabilities (CP), and 95% credible interval length (CIL) for θ^{CACE} . . . 55
3.7 The estimated probability of selecting a candidate model as the final model using DIC, based on simulation studies with 2000 replicates . . . 56

3.8 Performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets 57
4.1 Observed data and probabilities of a single study 65
4.2 Observed data and probabilities in study i of a meta-analysis with incomplete compliance 69
A1 Data in 10 randomized controlled trials of epidural analgesia in labor . . 108
A2 Summary of parameter estimates using Bayesian hierarchical models for 10 RCTs of epidural analgesia in labor 109
A3 Additional performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets 113

List of Figures

2.1 Forest plot of study-specific noncompliance rates in studies of epidural analgesia. The center of each square/circle and the horizontal lines represent the corresponding probability with 95% exact confidence interval. The diamonds indicate overall estimates of noncompliance rates using the bivariate generalized linear mixed effects model (BGLMM). This figure appears in color in the electronic version of this chapter. 15

2.2 Forest plot of CACE of epidural analgesia in labor on caesarean section. The center of each square and the horizontal lines represent the posterior medians and 95% equal tail credible intervals of θ_i^{CACE} for each individual study based on separate analyses. A diamond indicates the pooled estimate of θ^{CACE} and its 95% credible interval or confidence interval. The first overall estimate comes from the selected final model, Model IVa; the second one is from the REML estimator under a random-effect model using the two-step approach; the third one is the overall risk difference (RD) from the fixed-effect ITT analysis. This figure appears in color in the electronic version of this chapter. 20

3.1 Study-specific event rates vs. noncompliance rates in studies of epidural analgesia in labor. Coordinates of each dot are the posterior medians of the study-specific event rate and compliance rate. Horizontal lines represent the 95% CI of study-specific posterior cesarean section event rate. Vertical lines represent the 95% CI of study-specific posterior noncompliance rate. Dashed lines show results in the epidural analgesia arm, while solid lines mark show results in no/other analgesia group. The horizontal axis has a logarithmic scale. 32

3.2 Posterior distributions of θ^{CACE} of epidural analgesia in labor on cesarean section, the kernel smoothed density estimate from 100,000 Monte Carlo samples. 45

3.3 Forest plot of θ^{CACE} of epidural analgesia in labor on cesarean section. The center of each square and the horizontal lines represent the posterior median and 95% equal tail credible interval of θ_i^{CACE} for each study from the final model, Model Vc. The diamond indicates the pooled estimate of θ^{CACE} and its 95% credible interval. The symbol * indicates that the study has complete data on compliance status. With complete data, a solid horizontal line is used to represent the posterior 95% CI of θ_i^{CACE}, whereas a dashed line is used to show the CI for a study with incomplete compliance data. 48

3.4 Posterior of θ^{CACE} of the epidural analgesia in labor meta-analysis under the assumption that the missingness probability in the treatment arm p_{i1}^{mis} is linearly related to u_{i1} on the logit scale. Bold solid line: posterior median; fine solid lines: 95% equal-tail credible interval; fine dashed lines: 95% highest posterior density credible interval. The fine dotted horizontal line is $\theta^{\text{CACE}} = 0$. 51

4.1 Noncompliance rates plot generated by the function `plot.noncomp()`. . 73

4.2 Trace plot for θ^{CACE} on the `epidrual_ic` dataset from the model `cace.meta.ic()`. 87

4.3 The kernel smoothed density for θ^{CACE} from the function `cace.meta.ic()` applied to the epidural analgesia in labor meta-analysis. 88

4.4 Auto-correlation plot of θ^{CACE} from the model `cace.meta.ic()` fit to the `epidrual_ic` dataset. 89
4.5 Forest plot of study-specific θ^{CACE} from the model `cace.meta.ic()` with full random effects fit to the `epidural_ic` dataset. 91

Chapter 1

Introduction

1.1 Noncompliance and the causal effect in a single study

Well-conducted randomized controlled trials (RCTs) are considered the hallmark of evidence-based medicine and the gold standard for evaluating efficacy in clinical research (Holland, 1986). Yet in human randomized experiments, noncompliance with assigned treatments is common and may induce an information bias in estimating the average treatment effect (Frangakis and Rubin, 1999). One potential solution to this problem is intention-to-treat (ITT) analysis. It ignores noncompliance, withdrawal, and all post-randomization variables, and maintains prognostic balance based on the original random treatment allocation, but the estimated treatment effect is generally conservative (Freedman, 1990). Alternative analytic approaches are possible, including per-protocol analysis, which excludes noncompliant individuals, or as-treated analysis, which groups individuals by the treatment actually received. However, these approaches can break randomization and have a high potential for bias in either direction (Little and Yau, 1998).

Important statistical methods have been developed for analyzing RCT outcomes in the presence of noncompliance. Two frequently used estimands of treatment efficacy have been considered, the average treatment effect (ATE) (Little et al., 2009), and the complier average causal effect (CACE) (Yau and Little, 2001). The ATE is the

average efficacy across all patients and bounds on ATE have been derived for binary outcomes (Balke and Pearl, 1997). CACE analysis builds on the principal stratification framework (Frangakis and Rubin, 2002). The CACE is defined as the average difference in potential outcomes for the response in a subpopulation of subjects who would comply with their assigned treatment. The main challenge in CACE modeling is illustrated by considering treatment outcomes in the four latent classes defined by the intersection of treatment assignment and treatment compliance: *compliers* (for whom treatment received is the same as assigned), *never-takers* (who do not receive treatment, regardless of assignment), *always-takers* (who receive treatment regardless of assignment), and *defiers* (who always do the opposite of their treatment assignment).

The core insight of CACE analysis is that it retains the initial randomized assignment, so it overcomes the intrinsic bias associated with per-protocol or as-treated analysis. Under certain assumptions (see Section 2.1.1 for further discussion), we can derive an unbiased estimate of the difference in outcomes between compliers in the intervention group and those in the control group who would have engaged with treatment. Under some assumptions, CACE has been developed and applied in various settings to account for noncompliance. For example, Imbens and Rubin (1997) estimated CACE with the maximum-likelihood (ML) method using the EM algorithm and a Bayesian approach using the data augmentation algorithm. Little and Yau (1998) incorporated covariates in this framework via the ML-EM method. Cheng (2009) estimated CACE when the outcome is ordinal using the ML method. Methods to estimate causal effects under interference have also been developed recently (Hudgens and Halloran, 2008; Liu and Hudgens, 2014). However, most of those approaches only consider the setting of a single randomized clinical trial.

1.2 Causal effects in meta-analysis

To our knowledge, little attention has been paid to estimating causal effects from meta-analysis of multiple clinical trials accounting for noncompliance. Glasziou (1992) proposed a meta-analysis for "explanatory treatment effect" adjusting for compliance, based

on the linear model formulation of Newcombe (1988). Later, Baker and Kramer (2005) proposed ML estimates of "efficacy" when there is switching of interventions (which is similar to CACE) and briefly discussed its implications for meta-analysis using hypothetical sets of trials. Bannister-Tyrrell et al. (2015) presented fixed-effect meta-analysis adjusting for compliance in trials using an instrumental variable (IV) estimator. However, existing meta-regression methods cannot be used directly to derive CACE adjusting for noncompliance for the following reasons: 1) existing meta-regression can only be used to estimate the impact of fixed study-level covariates on study-specific relative effect size; 2) noncompliance is observed post-randomization and is considered to be a random variable; 3) noncompliance rates generally differ between treatment arms within a study and may affect the response to treatment. We address this gap by extending the CACE approach for ordinal outcomes from a single clinical trial setting to meta-analysis settings with multiple clinical trials. Compared to ITT analysis or a naive meta-regression approach using the averages of post-randomization variables (*e.g.*, each study's average noncompliance rate) as fixed covariates, this method allows us to appropriately account for the inherent heterogeneity in noncompliance rates between treatment arms and between studies.

1.3 Noncompliance with missing data in a single study

Besides noncompliance with treatment assignment, missing data also occur frequently in clinical trials and can affect their validity. Noncompliance occurs when some participants do not take or receive their assigned treatments, and data can be missing either for compliance status or for the outcome of interest. Missing outcome or compliance data happens when study investigators fail to collect those items on some subjects because of loss to follow-up or other reasons. Ignoring noncompliance or missing data may result in biased estimates of causal effects in the standard intention-to-treat analysis.

When a study has both noncompliance and missing outcome data, the CACE approach can still be used but further assumptions about the missing data mechanism are required. One commonly used assumption is "latent ignorability" (LI), which means that

the missing data are missing at random conditional on compliance status, *i.e.*, missingness has no residual dependence on the outcomes, given the observed data and the latent unobserved compliance classes. Under this assumption, several models that accommodate missing outcomes have been developed for inference about CACE (O'Malley and Normand, 2005; Peng et al., 2004). Also, Chen et al. (2009) discussed identifiability and estimation of CACE with missing outcome data under a nonignorabilility assumption, *i.e.*, when the missing data mechanism depends on the unobserved outcome. Analytical strategies for handling noncompliance are also increasingly used (Jo et al., 2010; Stuart et al., 2008), although not as widely as missing data methods.

1.4 Handling missing data and noncompliance in meta-analysis

Although inference for a clinical trial with noncompliance or missing data has been well studied, little attention has been paid to handling both missing data and noncompliance in meta-analysis. Meta-analysis, the statistical approach for synthesizing evidence from multiple sources, is gaining popularity in many fields due to the rapid growth of interest in comparative effectiveness research and evidence-based medicine (Egger et al., 2008). While multivariate and network meta-analysis (NMA) methods have been developed recently for meta-analyses of data consisting of multiple outcomes, multiple treatments, or multiple diagnostic tests (Lumley, 2002; Jackson et al., 2011; Zhang et al., 2014; Riley et al., 2017; Ma et al., 2018; Lian et al., 2018), important research gaps remain in meta-analysis in the area of causal inference. In particular, researchers have only recently started investigating causal effects in meta-analysis accounting for noncompliance (Baker and Kramer, 2005).

When noncompliance data are reported in each trial, intuitively one can first estimate CACE for each study, then combine these estimates using a meta-analytic method such as a fixed- or random-effect model to estimate the population-averaged CACE. We call this naive method a "two-step" approach. For ordinal or binary outcomes, we first

propose in Chapter 2 a Bayesian hierarchical model for estimating CACE in meta-analysis to account for noncompliance that is heterogeneous between studies, but this method is feasible only when the randomized allocation, actual taken treatment, and outcomes per treatment and allocation group of each study are recorded. However, existing meta-analysis methods cannot handle both noncompliance and missing data simultaneously. The two-step approach — which can be viewed as a special case of a model using only trials with complete noncompliance data — can thus be less efficient and potentially biased because trials without noncompliance data are excluded.

1.5 R packages to perform CACE analysis

An important step in promoting new statistical methods is to provide open-source user-friendly software. Several R packages are available to perform CACE analysis in a single study. For example, the **noncomplyR** package (Coggeshall, 2017) provides convenient functions for using Bayesian methods to perform inference on the CACE. The package **eefAnalytics** (Kasim et al., 2017) provides tools for exploratory CACE analysis of simple randomized trials, cluster randomized trials, and multi-site trials with focus on education trials. Besides the CACE analysis, another method quite commonly used to account for noncompliance is the instrumental variable (IV) method estimating the treatment effect with two-stage least squares (2SLS) regression (White, 1982); the R package **ivpack** (Jiang and Small, 2014) performs this type of analysis.

However, for analyzing multiple trials in the presence of noncompliance, little software is available for causal effect analysis and specifically for meta-analysis. When noncompliance data are reported in each trial, intuitively one could implement a two-step approach by first estimating CACE for each study and then combining the study-specific estimates using a fixed-effect or random effects model to estimate the population-averaged CACE. In Chapter 2 we propose a Bayesian hierarchical model to estimate the CACE in a meta-analysis of randomized trials where compliance may be heterogeneous between studies. This chapter is published in *Biometrics* (Zhou et al. (2019)). In addition, it is also common that noncompliance data is not available for some trials.

Simply excluding trials with incomplete noncompliance data from a meta-analysis can be inefficient and potentially biased. In Chapter 3 we propose an improved flexible Bayesian hierarchical CACE framework to account simultaneously for heterogeneous noncompliance and incomplete noncompliance data. The package **BayesCACE** focuses on providing user-friendly functions to estimate CACE in either a single study or meta-analysis using models based on Zhou et al. (2019) and on Chapter 3.

Chapter 2

A Bayesian Hierarchical Model Estimating CACE in Meta-analysis of Randomized Clinical Trials with Noncompliance

As introduced in Section 1.2, existing meta-regression methods cannot be used directly to estimate causal effects adjusting for noncompliance. We address this gap by extending the CACE approach for ordinal outcomes from a single clinical trial setting to meta-analysis settings with multiple clinical trials. Compared with ITT analysis or a naive meta-regression approach using the averages of post-randomization variables (*e.g.*, each study's average noncompliance rate) as fixed covariates, this method allows us to appropriately account for the inherent heterogeneity in noncompliance rates between treatment arms and between studies.

This chapter is organized as follows. Section 2.1 presents our Bayesian hierarchical modeling approach and explains the model assumptions. Section 2.2 gives a case study

reanalyzing a meta-analysis comparing epidural analgesia versus no or other analgesia in labor on the outcome of cesarean section, where noncompliance varies between studies. Section 2.3 presents a comprehensive simulation study evaluating the performance of our approach under a variety of conditions. Finally, Section 2.4 discusses our findings and implications for future developments.

2.1 Statistical Methods

2.1.1 Notation and assumptions

We consider I two-armed randomized trials. For the i^{th} trial, let N_i be the number of subjects, where N_{i0} are randomly assigned to the control group and N_{i1} are assigned to active treatment. Let $\boldsymbol{R}_i$ be the N_i-dimensional vector of randomization assignments for all subjects, with individual element $R_{ij} = r$ corresponding to whether subject j is assigned to treatment ($r = 1$) or control ($r = 0$). Let $\boldsymbol{T}_i^{\boldsymbol{r}}$ be the N_i-dimensional vector of *potential* treatments received under the randomization assignments $\boldsymbol{r}$ with individual element $T_{ij}^{\boldsymbol{r}}$, where $T_{ij}^{\boldsymbol{r}} = t \in \{0, 1\}$ according to whether subject j actually took the active treatment ($t = 1$) or placebo ($t = 0$) under $\boldsymbol{r}$. Let $\boldsymbol{Y}_i^{\boldsymbol{r},\boldsymbol{t}}$ be the vector of potential outcomes under randomization assignment $\boldsymbol{r}$ and treatment received $\boldsymbol{t}$, with individual element $Y_{ij}^{\boldsymbol{r},\boldsymbol{t}}$ for the j^{th} subject in the i^{th} trial. In this study, we consider only ordinal outcomes $Y_{ij}^{\boldsymbol{r},\boldsymbol{t}} = o \in \{1, \ldots, O\}$. Note that the sets of $\{Y_{ij}^{\boldsymbol{r},\boldsymbol{t}}\}$ and $\{T_{ij}^{\boldsymbol{r}}\}$ are *potential* outcomes and treatment-received status, where for each (i, j), only one of the possible values of each set can be observed, so we denote the observed response and received treatment as Y_{ij} and T_{ij} for the j^{th} subject in the i^{th} trial.

Following Imbens and Rubin (1997), we let C_{ij} be the latent compliance class of the j^{th} patient in the i^{th} trial, defined as follows:

$$C_{ij} = \begin{cases} 0, & \text{never-taker, if } (T_{ij}^0, T_{ij}^1) = (0, 0) \\ 1, & \text{complier, if } (T_{ij}^0, T_{ij}^1) = (0, 1) \\ 2, & \text{always-taker, if } (T_{ij}^0, T_{ij}^1) = (1, 1) \\ 3, & \text{defier, if } (T_{ij}^0, T_{ij}^1) = (1, 0). \end{cases}$$

In a two-arm trial, we can only observe one of T_{ij}^1 and T_{ij}^0 so a subject's compliance status C_{ij} is not an observable variable. It can only be partially identified based on treatment assignment and observed treatment received (see Table 2.1 columns R_{ij}, T_{ij}, and C_{ij}).

Table 2.1: Observed groups, latent compliance classes and outcome probabilities of trial i

R_{ij}	T_{ij}	C_{ij}	$Y_{ij} = o \in \{1, \ldots, O\}$	Count
0	0	0 (never-taker) or 1 (complier)	$M(n_{i00}, q_{io} = \frac{\pi_{ic} v_{io} + \pi_{in} s_{io}}{1 - \pi_{ia}})$	n_{i00o}
0	1	2 (always-taker) ~~or 3 (defier)~~	$M(n_{i01}, b_{io})$	n_{i01o}
1	0	0 (never-taker) ~~or 3 (defier)~~	$M(n_{i10}, s_{io})$	n_{i10o}
1	1	1 (complier) or 2 (always-taker)	$M(n_{i11}, p_{io} = \frac{\pi_{ic} u_{io} + \pi_{ia} b_{io}}{1 - \pi_{in}})$	n_{i11o}

Defiers are ruled out by the monotonicity assumption.

To identify treatment effects of interest, for each study we make assumptions identical to those listed in Angrist et al. (1996) and describe their implications below.

Assumption 1: Stable unit treatment value assumption (SUTVA) (Rubin, 1980). The outcome for a subject is unaffected by the particular assignments of treatments to the other individuals. That is, if $r = r'$ then $T_{ij}^{\boldsymbol{r}} = T_{ij}^{\boldsymbol{r'}}$; and if $r = r'$ and $t = t'$ then $Y_{ij}^{\boldsymbol{r},\boldsymbol{t}} = Y_{ij}^{\boldsymbol{r'},\boldsymbol{t'}}$.

Assumption 2: Random assignment to randomization groups. For all N_i subjects in the i^{th} trial, the treatment assignment $\boldsymbol{R}_i$ is random. This assumption implies that the proportion of each compliance class should be the same in the intervention and control groups.

Assumption 3: Exclusion restriction. For subject j in the i^{th} trial, $Y_{ij}^{r,t} = Y_{ij}^{r',t}$ for all r, r' and t, *i.e.*, the randomization assignment affects responses only through its effect on treatment received. With this assumption, *Assumption 1* can be restated: if $t = t'$ then $Y_{ij}^{\boldsymbol{r},\boldsymbol{t}} = Y_{ij}^{\boldsymbol{r'},\boldsymbol{t'}}$ for any r, r'. Therefore, for always-takers and never-takers, the distribution of outcomes does not depend on randomization group.

Assumption 4: Nonzero average causal effect of $\boldsymbol{R}_i$ on $\boldsymbol{T}_i$. For each trial, $E[T_{ij}^1 -$

$T_{ij}^0] \neq 0$. The fraction of subjects who receive each intervention varies by randomization group.

Assumption 5: Monotonicity. $P[T_{ij}^1 \geq T_{ij}^0] = 1$ for each trial. This means that no subject necessarily receives the treatment opposite to the randomized assignment, under assignment to either active treatment and control. This assumption rules out the existence of defiers, so it reduces the number of compliance types and permits a properly identified model (see Table 2.1).

2.1.2 Estimation and inference for the causal effect

For discrete outcomes, we extend the notation in Cheng (2009) and Baker (2011), and define the following parameters under the latent compliance model: 1) π_{ia}, the probability of being an always-taker in the i^{th} study; 2) π_{in}, the probability of being a never-taker in the i^{th} study, so the probability of being a complier in the i^{th} study, π_{ic}, is $1 - \pi_{ia} - \pi_{in}$; 3) u_{io}, the probability a complier randomized to the treatment group has outcome o in the i^{th} study; 4) v_{io}, the probability a complier randomized to the control group has outcome o in the i^{th} study; 5) s_{io}, the probability a never-taker has outcome o in the i^{th} study; and 6) b_{io}, the probability an always-taker has outcome o in the i^{th} study, where $\sum_{o=1}^{O} u_{io} = \sum_{o=1}^{O} v_{io} = \sum_{o=1}^{O} s_{io} = \sum_{o=1}^{O} b_{io} = 1$. The study-specific probabilities π_{ia}, π_{in}, u_{io}, v_{io}, s_{io}, and b_{io} may be assumed to be draws of their respective random effects, which follow some common distributions, discussed below.

Table 2.1 shows the distributions of observed n_{irt} in terms of the parameters for trial i, where $n_{irt} = \sum_j I(R_{ij} = r, T_{ij} = t)$ denotes the number of individuals in each observed group and $M(n_{irt}, \boldsymbol{x_{io}})$ denotes a multinomial distribution with n_{irt} subjects and multinomial probabilities $\{\boldsymbol{x_{io}}\}$. The observed count for each outcome o in group $\{j : R_{ij} = r, T_{ij} = t\}$ is n_{irto}, $o = 1, \ldots, O$. Although latent compliance classes cannot be fully identified based on randomization group (R_{ij}) and observed treatment-received behaviors (T_{ij}), the randomization assignment and the exclusion restriction assumptions imply that 1) the proportions of always-takers, never-takers and compliers in the control group are equal to those in the treatment group; 2) for never-takers and always-takers, the outcome distribution under control is the same as that under active treatment. They

further give $q_{io} = \frac{\pi_{ic}v_{io}+\pi_{in}s_{io}}{1-\pi_{ia}}$ and $p_{io} = \frac{\pi_{ic}u_{io}+\pi_{ia}b_{io}}{1-\pi_{in}}$ as probabilities corresponding to n_{i00o} and n_{i11o}, $o \in \{1, \ldots, O\}$ (shown in Table 2.1).

One causal effect of interest in many studies is the CACE discussed in Section 1.1. For an ordinal outcome $Y_{ij} = o \in \{1, \ldots, O\}$, suppose we can code scores $\{W_1, W_2, \ldots, W_O\}$ to reflect the real distances between categories. Then the CACE of the i^{th} trial is defined as

$$\theta_i^{\text{CACE}} = E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1) = \sum_o (W_o \times u_{io}) - \sum_o (W_o \times v_{io}). \tag{2.1}$$

The overall causal effect from the meta-analysis of I trials θ^{CACE} can be defined as $\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}})$. It is often unclear how to select scores $\{W_1, W_2, \ldots, W_O\}$ to reflect distances between categories. Equally spaced scores $\{1, 2, ..., O\}$ or their linear transforms are sensible in many cases and provide a reasonable compromise when there is no obvious choice (Agresti, 2003). Alternative scoring systems such as midranks are also possible. When uncertain about which scoring choice to use, a sensitivity analysis can consider different sensible choices to see how they affect the results.

2.1.3 Likelihood and the Bayesian hierarchical model

For the i^{th} trial, let λ_i denote the probability $P(R_{ij} = 1)$, which is usually known and can therefore be treated as fixed. Let $\boldsymbol{\beta}_i = (\pi_{ia}, \pi_{in}, \boldsymbol{s}_i, \boldsymbol{b}_i, \boldsymbol{u}_i, \boldsymbol{v}_i)$, where $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{i(O-1)})$, $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{i(O-1)})$, $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{i(O-1)})$, $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{i(O-1)})$, then the i^{th} trial's contribution to the likelihood is

$$L_i(\boldsymbol{\beta}_i) = \prod_j \prod_o \{(1-\lambda_i)(\pi_{ic}v_{io} + \pi_{in}s_{io})\}^{(1-R_{ij})(1-T_{ij})I(Y_{ij}=o)} \{(1-\lambda_i)\pi_{ia}b_{io}\}^{(1-R_{ij})T_{ij}I(Y_{ij}=o)}$$
$$\{\lambda_i\pi_{in}s_{io}\}^{R_{ij}(1-T_{ij})I(Y_{ij}=o)} \{\lambda_i(\pi_{ic}u_{io} + \pi_{ia}b_{io})\}^{R_{ij}T_{ij}I(Y_{ij}=o)}, \tag{2.2}$$

where $j = 1, \ldots, N_i, o = 1, \ldots, O$, $\pi_{ic} = 1 - \pi_{ia} - \pi_{in}$, the indicator function $I(Y_{ij} = o) = 1$ if $Y_{ij} = o$ and 0 otherwise, $\sum_o s_{io} = \sum_o b_{io} = \sum_o u_{io} = \sum_o v_{io} = 1$, and $0 \leq \pi_{ia}, \pi_{in}, \pi_{ic}, s_{io}, b_{io}, u_{io}, v_{io} \leq 1$.

The likelihood for all trials in a meta-analysis is $\mathcal{L}(\boldsymbol{\beta}) = \prod_i L_i(\boldsymbol{\beta}_i)$. Using the multivariate random effects meta-analysis framework (Jackson et al., 2011), one can specify

$\boldsymbol{\beta}_i \sim \boldsymbol{F}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{F}$ is some distribution function, $\boldsymbol{\beta}_0$ is the overall mean parameter and $\boldsymbol{\Sigma}_0$ is the variance-covariance matrix. In a Bayesian framework, one would also specify the prior distributions of $\boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_0$ as $f(\boldsymbol{\beta}_0)$ and $f(\boldsymbol{\Sigma}_0)$, respectively. The joint posterior distribution is then proportional to $\prod_i L_i(\boldsymbol{\beta}_i)\boldsymbol{F}(\boldsymbol{\beta}_i|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)f(\boldsymbol{\beta}_0)f(\boldsymbol{\Sigma}_0)$. The posterior distribution can be estimated by Markov chain Monte Carlo (MCMC) methods, specifically the Gibbs and Metropolis-Hastings sampling algorithms (Gelfand and Smith, 1990).

2.1.4 Accounting for heterogeneity through random effects models

Between-study heterogeneity commonly exists in a meta-analysis because studies usually differ in their subject recruitment methods, measurement techniques, study qualities, *etc.*. To account for potential between-study heterogeneity of the fractions in the compliance classes and of the outcome probabilities, we consider a random-effects model.

To guarantee the desired properties of the probabilities of being in each principal stratum in study i, *i.e.*, $\pi_{in} + \pi_{ia} + \pi_{ic} = 1$, and $0 \leq \pi_{in}, \pi_{ia}, \pi_{ic} \leq 1$, and to allow these probabilities to vary between studies, we assume $\pi_{in} = \frac{exp(n_i)}{1+exp(n_i)+exp(a_i)}$, $\pi_{ia} = \frac{exp(a_i)}{1+exp(n_i)+exp(a_i)}$, where $n_i = \alpha_n + \delta_{in}$, $a_i = \alpha_a + \delta_{ia}$, and $(\delta_{in}, \delta_{ia})^T \sim N(0, \boldsymbol{\Sigma}_{ps})$. To allow correlation between n_i and a_i, the variance-covariance matrix is $\boldsymbol{\Sigma}_{ps} = \begin{pmatrix} \sigma_n^2 & \rho\sigma_n\sigma_a \\ \rho\sigma_n\sigma_a & \sigma_a^2 \end{pmatrix}$.

To be consistent with the case study presented in Section 2.2, we describe the rest of the model for the binary case, where $o \in \{0, 1\}$, *i.e.*, $s_{i0} + s_{i1} = b_{i0} + b_{i1} = u_{i0} + u_{i1} = v_{i0} + v_{i1} = 1$ for the i^{th} trial, so the vector parameters of $\boldsymbol{s}_i, \boldsymbol{b}_i, \boldsymbol{u}_i, \boldsymbol{v}_i$ become scalars $s_{i1}, b_{i1}, u_{i1}, v_{i1}$. Specifically, we assume a normal distribution for each response probability s_{i1}, b_{i1}, u_{i1}, and v_{i1} on the transformed scales, and allow heterogeneity between studies: $g(s_{i1}) = \alpha_s + \delta_{is}$, $g(b_{i1}) = \alpha_b + \delta_{ib}$, $g(u_{i1}) = \alpha_u + \delta_{iu}$, $g(v_{i1}) = \alpha_v + \delta_{iv}$, where $g(\cdot)$ is a link function such as the logit or probit. We then assume the probabilities of having an outcome are independent across principal strata, so $\delta_{is} \sim N(0, \sigma_s^2)$, $\delta_{ib} \sim N(0, \sigma_b^2)$, $\delta_{iu} \sim N(0, \sigma_u^2)$, $\delta_{iv} \sim N(0, \sigma_v^2)$. To extend this model to more general cases with outcomes taking more than two values, one could fit a generalized logit model, *e.g.*, the baseline-category logit model for nominal responses or the cumulative logit model for ordinal responses (Agresti, 2003).

2.1.5 Implementation

We used a Bayesian hierarchical model (Gelman et al., 2013), fit using JAGS. To avoid over-fitting the data with an excess of random effects, we used a forward selection procedure on the random effects. Specifically, at each forward step, we added the random-effect component that provided the largest improvement in the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Other model-selection approaches are easily substituted for these choices, *e.g.*, using a different model-selection criterion (Celeux et al., 2006), or a different search strategy.

Computation used Markov chain Monte Carlo (MCMC) in JAGS, interfacing with R using the package "rjags". Each fit consisted of three independent chains initiated with parameter values drawn randomly from their prior distributions. Each chain was allowed 10,000 iterations of burn-in, and the subsequent 100,000 iterations were collected as posterior samples. Estimates are presented as the posterior mean with a 95% credibility interval. Convergence of Markov chains was assessed using trace plots, sample autocorrelation, and the Gelman and Rubin convergence statistic (Gelman and Rubin, 1992). We also used the R package "mcmcse" (Flegal et al., 2017) to calculate the effective sample size and to make sure the chains gave adequately precise posterior estimates.

2.2 Case Study

2.2.1 Data and descriptive analysis

We reanalyzed the data from a meta-analysis of the impact of epidural analgesia in labor compared with no or other analgesia, on the outcome of cesarean section (Bannister-Tyrrell et al., 2015). Epidural analgesia in labor is a highly effective method of labor pain relief, but there is a controversy over whether epidural analgesia in labor is associated with an increased risk of cesarean delivery. However, noncompliance is common in RCTs in obstetrics: it is unethical to deny laboring women their preferred mode of labor analgesia because trial entry and randomization occurs before the experience of

labor pain. Thus the ITT result is not a good estimate of the consequence of receiving epidural analgesia due to noncompliance, and it motivates us to investigate the causal effect. Data were recorded on randomized treatment assignment (1=Epidural analgesia, 0=No/other analgesia in labor), treatment actually received (1=Epidural analgesia, 0=No/other analgesia in labor), and frequency of cesarean section (1=Yes, 0=No) by compliance with allocated intervention. 10 RCTs had complete data available on the number of cesarean sections in compliant and noncompliant participants, which could be meta-analyzed.

In these 10 trials, 1,684 women were randomly assigned to epidural analgesia, with 130 cesarean deliveries, and 1,732 were assigned to no/other analgesia with 115 cesarean deliveries. Unlike the ITT and IV analysis used by Bannister-Tyrrell et al. (2015), in which the data from Nikkola et al. (1997) were excluded because it had no cesarean section events, we included this trial because our Bayesian method accommodates trials with zero events.

Figure 2.1 displays the noncompliance rates of these studies, where $\mathrm{P}(T = 1|R = 0)$ is the fraction of participants who were randomized to no/other analgesia in labor but actually received epidural analgesia, and $\mathrm{P}(T = 0|R = 1)$ is the fraction of participants who did not receive epidural analgesia among those who were assigned to it. Because several arms had 0 events, the confidence intervals were calculated using an exact method, the Clopper-Pearson exact interval (Clopper and Pearson, 1934). We adopted a bivariate generalized linear mixed effects model (BGLMM) (Chu and Cole, 2006) to model the two overall noncompliance rates. The BGLMM assumes a bivariate normal distribution for the logits of the probabilities in the two groups: $\mathrm{logit}(p_{1i}) = \mu_1 + \eta_{1i},\ \mathrm{logit}(p_{0i}) = \mu_0 + \eta_{0i},\ (\eta_{1i}, \eta_{0i})^T \sim MVN(\mathbf{0}, \mathbf{\Sigma}_\eta)$, where $p_{0i} = \mathrm{P}(T = 1|R = 0)$ and $p_{1i} = \mathrm{P}(T = 0|R = 1)$. As shown in Figure 2.1, the noncompliance rates varied substantially across studies. The index of heterogeneity I^2 (Higgins and Thompson, 2002) was 98.48% for $\mathrm{P}(T = 1|R = 0)$ and 98.07% for $\mathrm{P}(T = 0|R = 1)$. The high heterogeneity motivated us to apply Bayesian hierarchical models to estimate CACE accounting for between-study heterogeneity.

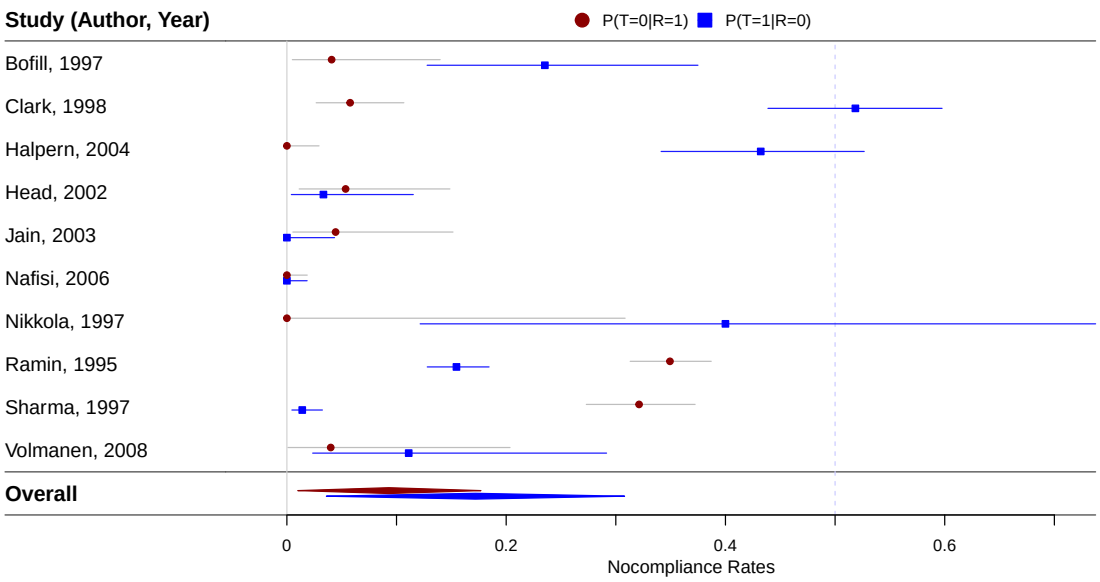

Figure 2.1: Forest plot of study-specific noncompliance rates in studies of epidural analgesia. The center of each square/circle and the horizontal lines represent the corresponding probability with 95% exact confidence interval. The diamonds indicate overall estimates of noncompliance rates using the bivariate generalized linear mixed effects model (BGLMM). This figure appears in color in the electronic version of this chapter.

In this case study, the SUTVA assumption holds because a participant's outcome did not depend on the other participants' treatments. The exclusion restriction assumption holds because a participant's randomized assignment affected the cesarean section outcome only through the analgesia that was actually received. The other three assumptions are also reasonable based on the randomized clinical trial design.

2.2.2 Likelihood and priors

As discussed in Section 2.1.4, for a binary outcome ($o \in \{0, 1\}$) the parameter vector is reduced to $\boldsymbol{\beta}_i = (\pi_{ia}, \pi_{in}, s_{i1}, b_{i1}, u_{i1}, v_{i1})$. From Table 2.1, the probability of each observed n_{irto} is determined by the above parameters. Therefore, the i^{th} trial's contribution to the likelihood is

$$L_i(\boldsymbol{\beta}_i) = [(1-\lambda_i)\{\pi_{ic}(1-v_{i1}) + \pi_{in}(1-s_{i1})\}]^{n_{i000}}\{(1-\lambda_i)(\pi_{ic}v_{i1} + \pi_{in}s_{i1})\}^{n_{i001}}$$

$$\{(1-\lambda_i)\pi_{ia}(1-b_{i1})\}^{n_{i010}}\{(1-\lambda_i)\pi_{ia}b_{i1}\}^{n_{i011}}\{\lambda_i\pi_{in}(1-s_{i1})\}^{n_{i100}}$$

$$\{\lambda_i\pi_{in}s_{i1}\}^{n_{i101}}[\lambda_i\{(\pi_{ic}(1-u_{i1}) + \pi_{ia}(1-b_{i1})\}]^{n_{i110}}\{\lambda_i(\pi_{ic}u_{i1} + \pi_{ia}b_{i1})\}^{n_{i111}} \qquad (2.3)$$

and the likelihood for all trials in the meta-analysis is $\mathcal{L}(\boldsymbol{\beta}) = \prod_i L_i(\boldsymbol{\beta}_i)$.

Using equally spaced scores ($W_0 = 0, W_1 = 1$), the CACE for study i is $\theta_i^{\text{CACE}} = u_{i1} - v_{i1}$, the response rate difference in compliers. Using the model introduced in Section 2.1.4, the medians of u_{i1} and v_{i1} can be estimated as $M(u_{i1}) = g^{-1}(\alpha_u)$ and $M(v_{i1}) = g^{-1}(\alpha_v)$ respectively, while the means of u_{i1} and v_{i1} can be estimated as $E(u_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t)\sigma_u^{-1}\phi(\frac{t}{\sigma_u})dt$ and $E(v_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_v + t)\sigma_v^{-1}\phi(\frac{t}{\sigma_v})dt$ respectively, where $\phi(\cdot)$ is the standard Gaussian density. Thus $\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}}) = E(u_{i1}) - E(v_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t)\sigma_u^{-1}\phi(\frac{t}{\sigma_u})dt - \int_{-\infty}^{+\infty} g^{-1}(\alpha_v + t)\sigma_v^{-1}\phi(\frac{t}{\sigma_v})dt$. Although $E(\theta_i^{\text{CACE}})$ involves an integral, the probit-link random effects model implies closed-form formulas $E(u_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}})$ and $E(v_{i1}) = \Phi(\frac{\alpha_v}{\sqrt{1+\sigma_v^2}})$, and the logit-link random effects model has a well-established approximation, $E(u_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_u}{\sqrt{1+C^2\sigma_u^2}})$ and $E(v_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_v}{\sqrt{1+C^2\sigma_v^2}})$, where $C = \frac{16\sqrt{3}}{15\pi}$ (Zeger et al., 1988).

We selected proper but diffuse prior distributions for the hyper-parameters because non-informative prior distributions can lead to inaccurate posterior estimates (Natarajan and McCulloch, 1998). For fixed effects, vague priors were assigned. In particular, α_n and α_a both follow $N(0, 2.5^2)$, so under the simplest situation (a fixed effects model), a 95% prior probability interval for any of the probabilities $\pi_{in}, \pi_{ia}, \pi_{ic}$ would range from about 0.001 to 0.91; and $\alpha_s, \alpha_b, \alpha_u, \alpha_v$ all follow $N(0, 2^2)$, which implies a 95% interval for the probabilities $s_{i1}, b_{i1}, u_{i1}, v_{i1}$ ranging from about 0.01 to 0.98. The hyper-priors for the precision parameters $\sigma_s^{-2}, \sigma_b^{-2}, \sigma_u^{-2}, \sigma_v^{-2}$ were assumed to be $Gamma(2, 2)$, which implies a 95% interval $(0.36, 8.36)$ for the variance parameters, allowing moderate heterogeneity for the response probabilities. The prior for the precision matrix $\boldsymbol{\Sigma}_{ps}^{-1}$ was Wishart, *i.e.*, $W(I, 3)$, where I is the identity matrix. In the reduced model with one of σ_n^2, σ_a^2 set to 0, the prior of the other precision parameter was still assumed to be

$Gamma(2, 2)$, which gave moderate heterogeneity for latent principal strata probabilities.

2.2.3 Model selection and results

We adopted the forward selection procedure introduced in Section 2.1.5 to search for the best-fitting model. The models specified in Section 2.1.4 have 6 potential random effects in total: δ_{in}, δ_{ia}, δ_{is}, δ_{ib}, δ_{iu} and δ_{iv}. In each step of the forward-selection procedure, we added the random-effect component that provided the largest improvement in DIC. Table 2.2 presents DIC, DIC improvement, and the effective number of parameters (p_D) for each model fit to the 10 studies. Starting with the simplest model with no random effects (Model I), the largest improvement was obtained by allowing the probability of being an always-taker to be a random effect, *i.e.*, adding δ_{ia} (Model IIf). In the second step, adding a random effect for the probability of being a never-taker δ_{in} (Model IIIe) decreased DIC the most. This revealed an important characteristic of this meta-analysis: the studies might vary considerably in their recruitment criteria, study procedures, beliefs of the local PIs, *etc.*, resulting in different properties of the latent compliance classes. The next forward step produced a meaningful improvement by including a random effect for the cesarean section rate of a never-taker, δ_{is} (Model IVa). The resulting improvement was modest compared to adding random effects δ_{ia} and δ_{in}. It is difficult to say what constitutes an important difference in DIC; we follow Lunn et al. (2012) in considering that a reduction of less than 5 is not a substantial improvement. In the last step, DIC was reduced only by 3.6 compared to Model IVa, so the final model included random effects δ_{ia}, δ_{in}, and δ_{is}.

Table 2.2: Selection of random effects using a forward selection procedure on RCTs of epidural analgesia in labor

Random effects models	DIC	DIC improvement	p_D
Model **I** (None)	**917.4**	N/A	6.0
Model IIa (δ_{is})	896.8	20.6	21.3
Model IIb (δ_{ib})	918.3	-0.9	16.1
Model IIc (δ_{iu})	915.0	2.4	12.4
Model IId (δ_{iv})	907.4	10.0	14.1
Model IIe (δ_{in})	577.1	340.3	13.6
Model **IIf** (δ_{ia})	**537.2**	**380.2**	15.7
Model IIIa (δ_{ia}, δ_{is})	514.9	22.3	29.3
Model IIIb (δ_{ia}, δ_{ib})	538.6	-1.4	19.5
Model IIIc (δ_{ia}, δ_{iu})	531.9	5.3	21.1
Model IIId (δ_{ia}, δ_{iv})	526.3	10.9	22.9
Model **IIIe** (δ_{ia}, δ_{in})	**265.7**	**271.5**	21.4
Model **IVa** ($\delta_{ia}, \delta_{in}, \delta_{is}$)	**246.5**	**25.0**	27.3
Model IVb ($\delta_{ia}, \delta_{in}, \delta_{ib}$)	266.5	5.0	24.9
Model IVc ($\delta_{ia}, \delta_{in}, \delta_{iu}$)	262.1	9.4	28.2
Model IVd ($\delta_{ia}, \delta_{in}, \delta_{iv}$)	267.1	4.4	29.3
Model IVe ($\delta_{ia}, \delta_{in}, \rho$)	265.7	5.8	22.1
Model Va ($\delta_{ia}, \delta_{in}, \delta_{is}, \rho$)	246.2	0.3	27.6
Model Vb ($\delta_{ia}, \delta_{in}, \delta_{is}, \delta_{ib}$)	247.2	-0.7	30.7
Model Vc ($\delta_{ia}, \delta_{in}, \delta_{is}, \delta_{iu}$)	242.9	3.6	34.0
Model Vd ($\delta_{ia}, \delta_{in}, \delta_{is}, \delta_{iv}$)	253.0	-6.5	34.4

Bold-face cells are models selected using the forward selection procedure. DIC improvement is the reduction in DIC compared to the previous step's lowest DIC. The study-specific component corresponding to each random effect: δ_{is} never-taker response, δ_{ib} always-taker response, δ_{iu} treated complier response, δ_{iv} control complier response, δ_{in} never-taker probability, δ_{ia} always-taker probability; ρ is the correlation between δ_{ia} and δ_{in}.

Figure 2.2 is a forest plot of θ_i^{CACE} for each study individually, using the Bayesian method with the foregoing priors. The first overall estimate of θ^{CACE} is from the final model (Model IVa). The second estimate is the REML (restricted maximum likelihood) estimate calculated using the individual posterior means and standard errors of the θ_i^{CACE}. We call this method the "two-step" approach because it first analyzes each study separately, then combines the individual estimates to give a pooled eastimate of θ^{CACE}. The estimated θ_i^{CACE} varied from negative to positive in individual studies, but the 95% credible intervals all cover zero. Among overall estimates, the two-step approach gave a wider interval than our proposed method, suggesting it may not be efficient because it required the whole set of parameters to be estimated for each study, leading to a larger total number of parameters.

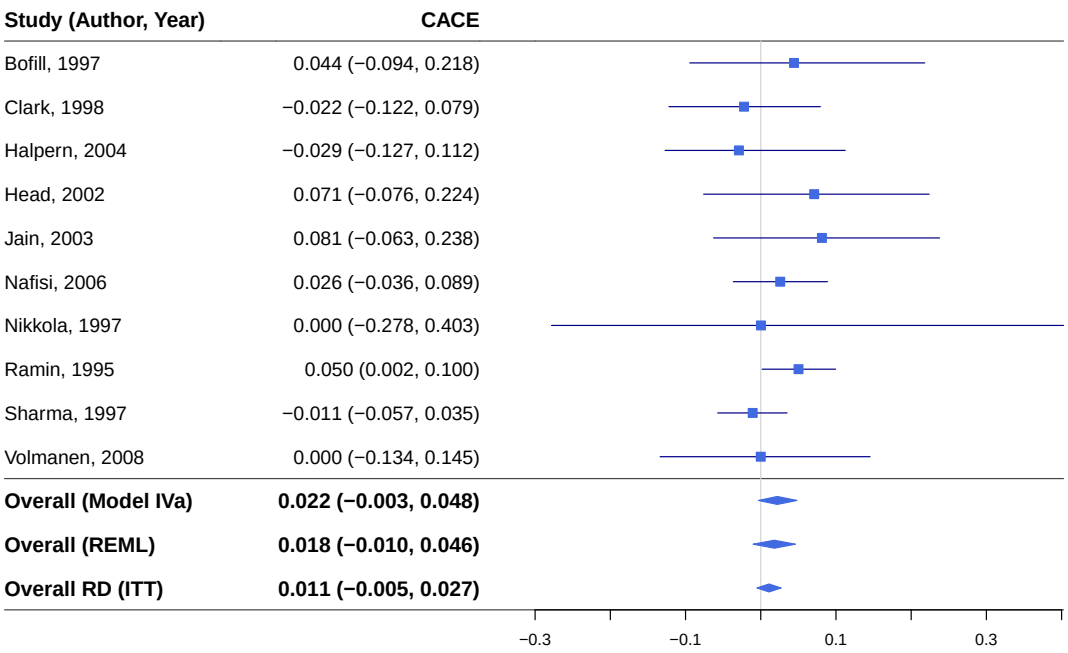

Figure 2.2: Forest plot of CACE of epidural analgesia in labor on caesarean section. The center of each square and the horizontal lines represent the posterior medians and 95% equal tail credible intervals of θ_i^{CACE} for each individual study based on separate analyses. A diamond indicates the pooled estimate of θ^{CACE} and its 95% credible interval or confidence interval. The first overall estimate comes from the selected final model, Model IVa; the second one is from the REML estimator under a random-effect model using the two-step approach; the third one is the overall risk difference (RD) from the fixed-effect ITT analysis. This figure appears in color in the electronic version of this chapter.

The ITT fixed-effect meta-analysis gave an estimated risk difference of 0.011 with 95% confidence interval (-0.005, 0.027); the overall θ^{CACE} from our final model, 0.022 $(-0.003, 0.048)$, was not significant either, though the point estimate was about twice of that from the ITT analysis in the same direction. Thus we conclude that the potential

dilution of the ITT estimated treatment effect notwithstanding, epidural analgesia in labor does not affect the risk of cesarean section in the view of causal interpretation.

2.3 Simulation Studies

To evaluate the performance of our modeling approach and to study the impact of misspecification of random effects, we performed four sets of simulations. For ease of presentation and interpretation, we generated data with, at most, random effects only for the latent compliance class probabilities and the response probabilities of treated compliers, $(\delta_{in}, \delta_{iu})$, then fit models with up to three random effects $(\delta_{in}, \delta_{iu}, \delta_{iv})$. Specifically, data were simulated from the following four models: 1) no random effects; 2) random effect for latent compliance class probabilities δ_{in}; 3) random effect for the response probabilities of compliers in the treatment group δ_{iu}; and 4) random effects for both δ_{in} and δ_{iu}. The simulations are realistic scenarios that researchers are likely to encounter, such as those in the case studies.

We simulated 2000 meta-analysis datasets in each setting. For each dataset, 20 studies were simulated, each with 350 observations, roughly matching the sample sizes in our case study. In each simulated study, the allocation ratio was 1:1 ($\lambda_i = 0.5$). True values were set to $(\alpha_n, \alpha_a, \alpha_s, \alpha_b, \alpha_u, \alpha_v) = (-0.4, -0.6, 0.5, -0.5, -0.5, 0.5)$, giving true values $\pi_{in} = 0.302$, $\pi_{ia} = 0.247$ and $\theta_i^{\text{CACE}} = -0.383$ in the absence of random effects. In the presence of random effects, the variances of δ_{in}, δ_{iu} were set to 0.5^2 so the true θ^{CACE} was approximately -0.364 when δ_{iu} was present.

For each simulated dataset, we fit models with: 1) no random effects; 2) one random effect, for δ_{in}, δ_{iu}, or δ_{iv}; 3) two random effects, for $[\delta_{in}, \delta_{iu}]$, $[\delta_{in}, \delta_{iv}]$, or $[\delta_{iu}, \delta_{iv}]$; and 4) three random effects, for $[\delta_{in}, \delta_{iu}, \delta_{iv}]$. Each analysis used JAGS with forward selection using DIC. Table 2.3 summarizes the estimated probability of selecting each candidate model as the "best" model in each simulation setting. DIC identified the true random effects model with probability over 0.95, i.e., our procedure usually selected the correct random effects and thus usually accounted properly for uncertainty in the estimates.

Table 2.3: Estimated probability of selecting each candidate model using DIC*, based on 2000 simulated datasets

True Random Effects Model	Selected Random Effects Model							
	None	δ_{in}	δ_{iu}	δ_{iv}	δ_{in}, δ_{iu}	δ_{in}, δ_{iv}	δ_{iu}, δ_{iv}	$\delta_{in}, \delta_{iu}, \delta_{iv}$
None	**95.05**	1.60	1.85	1.45	0	0.05	0	0
δ_{in}	0	**96.25**	0	0	1.85	1.90	0	0
δ_{iu}	0	0	**96.65**	0	1.25	0	2.00	0.10
δ_{in}, δ_{iu}	0	0	0	0	**98.25**	0	0	1.75

Bold-face cells give the estimated probability of identifying the correct model. The numbers have been multiplied by 100 for presentation.

*DIC: deviance information criterion.

Table 2.4: Performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets

True Random Effects Model		Selected Random Effects Model							
		None	δ_{in}	δ_{iu}	δ_{iv}	δ_{in}, δ_{iu}	δ_{in}, δ_{iv}	δ_{iu}, δ_{iv}	δ_{in}, δ_{iu}, δ_{iv}
None	Mean	**-0.383**	-0.383	-0.373	-0.374	-0.372	-0.374	-0.363	-0.363
	Bias	**0.000**	0.000	0.010	0.009	0.011	0.009	0.020	0.020
	95% CIL*	**0.100**	0.099	0.171	0.171	0.170	0.172	0.220	0.221
	95% CICp**	**0.947**	0.948	0.999	0.999	0.999	0.999	1.000	1.000
δ_{in}	Mean	-0.383	**-0.382**	-0.372	-0.374	-0.372	-0.373	-0.364	-0.363
	Bias	0.000	**0.001**	0.011	0.009	0.011	0.010	0.019	0.020
	95% CIL	0.101	**0.098**	0.172	0.173	0.170	0.175	0.222	0.223
	95% CICp	0.946	**0.947**	0.998	0.999	0.998	0.999	1.000	1.000
δ_{iu}	Mean	-0.365	-0.364	**-0.356**	-0.355	-0.355	-0.355	-0.346	-0.346
	Bias	-0.001	0.000	**0.008**	0.009	0.009	0.009	0.018	0.018
	95% CIL	0.100	0.099	**0.204**	0.171	0.204	0.172	0.247	0.248
	95% CICp	0.755	0.759	**0.986**	0.944	0.986	0.946	0.995	0.996
δ_{in}, δ_{iu}	Mean	-0.365	-0.365	-0.357	-0.356	**-0.356**	-0.355	-0.348	-0.347
	Bias	-0.001	-0.001	0.008	0.008	**0.008**	0.009	0.016	0.017
	95% CIL	0.101	0.099	0.205	0.173	**0.204**	0.175	0.249	0.250
	95% CICp	0.763	0.753	0.986	0.951	**0.987**	0.951	0.997	0.996

Bold-face cells are the correct model.

*95% CIL: 95% equal-tail credible interval length.

**95% CICp: 95% credible interval coverage probability.

Table 2.4 shows the estimated mean, bias, 95% credible interval length, and coverage probability for θ^{CACE} under each model. We present results only for θ^{CACE} because of space limitations. In general, the posterior standard deviation becomes larger as more random effects are included. Over-fitting (including a random effect when it is absent) tends to give longer 95% credible intervals. As for the coverage probabilities, 1) under the correct model or when over-fitting occurs, the coverage probabilities are close to or greater than the nominal 0.95; and 2) when under-fitting occurs (omitting a random effect when it is present), failure to include the random effect for the study-specific probability of latent compliance groups (δ_{in}) does not substantially affect the coverage probability, but failure to include the random effect for the response rate of a treated

complier (δ_{iu}) reduces coverage for θ^{CACE} notably.

2.4 Discussion

To estimate complier average causal effects in meta-analysis of RCTs with noncompliance, we proposed a Bayesian hierarchical model that accounts for between-study heterogeneity and applied it to a meta-analysis of epidural analgesia trials. We also performed simulation studies to evaluate the performance of our approach and the impact of misspecification of random effects. To the best of our knowledge, this is the first meta-analysis of RCTs estimating CACE accounting for noncompliance.

In the case study we compared the estimate of θ^{CACE} from the "two-step approach" with the one from our proposed final model. The two-step approach can be a valid alternative to our proposed random-effect model. However, this approach requires the whole set of parameters to be estimated independently for each study. Thus, the total number of effective parameters may be larger, making it potentially not as efficient as the one-step approach. Also, after computing θ_i^{CACE} and $\text{SE}(\theta_i^{\text{CACE}})$ from posterior samples (which depends on large-sample theory), various methods could be used to account for the heterogeneity besides the REML estimator we used, *e.g.*, the DerSimonian-Laird, Hedges, or Hunter-Schmidt estimators (Viechtbauer, 2005), which may complicate the estimation.

The control procedure in this case study is defined as "nonepidural or no analgesia in labor", so the control treatment varies between studies, making it possible that the corresponding population of each latent class is not identical across studies. Because θ_i^{CACE} in Figure 2.2 was estimated from separate analysis only using data from study i, the study-specific θ_i^{CACE} could be interpreted slightly different depending on what the particular control procedure for that study was. However, the different subpopulations of compliers in individual studies are considered subsets of the general "compliers" population defined for the meta-analysis comparing epidural analgesia in labor with nonepidural or no analgesia. The estimand θ^{CACE} is thus still sufficiently well-defined and its interpretation is not affected.

Note that the CACE is defined under the principal stratification framework assuming SUTVA, random assignment, the exclusion restriction, $E[T_{ij}^1 - T_{ij}^0] \neq 0$, and monotonicity. In our case study, the assumptions are plausible as discussed in Section 2.2.1. In other cases, some assumptions may be violated; accounting for these violations will increase the complexity of defining and estimating causal effects. Moreover, the CACE analysis typically assumes that compliance status is reported without error, but this might be invalid for self-reported compliance. Therefore, further research is needed on the consequences of relaxing key assumptions and incorporating noncompliance measurement errors. Some extensions have been developed for a single randomized trial, *e.g.*, estimating causal effects in the presence of interference (Liu and Hudgens, 2014), noncompliance measured with error (Imai and Yamamoto, 2010), and incorporating baseline covariates (Roy et al., 2007). Extensions to meta-analysis of RCTs incorporating those issues await further development.

Also, as a reviewer noted, in our epidural analgesia case study, compliance may be strongly determined by external factors such as the severity of pain. Pain severity is another post-randomization variable that is determined after randomization and may relate to both the actual taken treatment and the outcome. Randomization should have made the pain severity balanced between treatment arms. However, pain may influence the behavior of compliance differently in the treatment and control arms: in the epidural analgesia group, severe labor pain may reduce the rate of noncompliance, while in the control group, severe labor pain possibly increases the rate of noncompliance. Other possible post-randomization variables could include the length of labor. However, it is hard to measure pain severity and for our example meta-analysis, we do not have data on that or other post-randomization variables. Methods for handling such post-randomization variables deserve further investigation.

2.5 Supporting Information

Data analyzed in Section 2.2, parameter estimates from the fixed effects model (Model I) and the final model (Model IVa), sample R JAGS code implementing Model IVa

for the case study, sensitivity analysis on prior distributions, and additional simulation scenarios are available in the Appendix at the end of the thesis.

Chapter 3

A Bayesian Hierarchical CACE Model Accounting for Incomplete Noncompliance Data in Meta-analysis

The present chapter's main purpose is to develop a flexible statistical framework to use noncompliance data that is both heterogeneous and incomplete across studies, in a meta-analysis of RCTs with ordinal or binary outcomes. The idea is motivated by a meta-analysis, conducted by Bannister-Tyrrell et al. (2015), of the effect of epidural analgesia in labor on the occurrence of cesarean section, in which only 9 of 27 studies were included because only 10 studies had full compliance data and 1 other study was excluded because it had zero cesarean section events. Our proposed Bayesian hierarchical framework can include studies that do not report noncompliance data and studies with zero events.

This rest of this chapter is organized as follows. Section 3.1 describes the motivating case study of epidural analgesia, in which noncompliance varies between studies and compliance status was missing for 17 of 27 studies. Section 3.2 first presents the assumptions for estimating the causal effect and for missingness, then describes our

Bayesian hierarchical modeling approach and describes how we obtained posterior distributions for the overall and study-specific CACEs. Section 3.3 applies the model to the epidural analgesia case study using a particular approach to model selection and conducts sensitivity analysis to the missing data assumptions. Section 3.4 presents simulation studies evaluating the performance of our approach under a variety of conditions. Finally, Section 3.5 discusses our findings and potential extensions in future work.

3.1 A Motivating Study

3.1.1 Data Sources

Epidural analgesia in labor is a highly effective method of labor pain relief, but it remains controversial whether epidural analgesia in labor increases the risk of cesarean section delivery. Nonetheless, good evidence to support or refute this association is still limited, mainly because randomized controlled trials in obstetrics often have high rates of noncompliance. In this setting, the consequences of receiving epidural analgesia are more important to clinicians and patients than the impact of being assigned to epidural analgesia, thus the intention-to-treat (ITT) result is not a good estimate of the consequence of receiving epidural analgesia due to noncompliance. Bannister-Tyrrell et al. (2015) conducted an exploratory meta-analysis of the association between epidural analgesia in labor and cesarean section by using the 9 trials, out of the 27 RCTs included in the systematic review, that have full compliance data with non-zero events.

Data were recorded on treatment assignment r ($r = 1$ for epidural analgesia, $r = 0$ for no/other analgesia in labor), actual received intervention t ($t = 1$ for epidural analgesia, $t = 0$ for no/other analgesia in labor), and frequency of cesarean section o ($o = 1$ for yes, $o = 0$ for no) by compliance with assigned intervention, where noncompliance describes participants who were randomly assigned to receive epidural analgesia in labor but who in fact received either another or no analgesia, or who were assigned to the control group but ultimately received epidural analgesia in labor. Then for study i ($i = 1, 2, ..., I$), the count N_{irto} denotes the number of patients in randomization group r who received intervention t and had outcome o.

Table 3.1: Data from randomized controlled trials of epidural analgesia in labor

Study	Author, Year	Complete data								Missing data			
		Allocated control				Allocated epidural				Allocated control		Allocated epidural	
		Received Control		Received epidural		Received Control		Received epidural					
		Cesarean – N_{i000}	+ N_{i001}	Cesarean – N_{i010}	+ N_{i011}	Cesarean – N_{i100}	+ N_{i101}	Cesarean – N_{i110}	+ N_{i111}	Cesarean – N_{i0*0}	+ N_{i0*1}	Cesarean – N_{i1*0}	+ N_{i1*1}
1	Bofill, 1997 *	37	2	11	1	2	0	42	5	0	0	0	0
2	Clark, 1998 *	72	6	68	16	7	2	134	13	0	0	0	0
3	Dickinson, 2002	0	0	0	0	0	0	0	0	428	71	408	85
4	Evron, 2008	40	4	0	0	0	0	0	0	0	0	129	19
5	El Kerdawy, 2010	0	0	0	0	0	0	0	0	12	3	11	4
6	Gambling, 1998	0	0	0	0	206	10	371	29	573	34	0	0
7	Grandjean, 1979	0	0	0	0	0	0	0	0	59	1	30	0
8	Halpern, 2004 *	62	5	44	7	0	0	112	12	0	0	0	0
9	Head, 2002 *	51	7	2	0	3	0	43	10	0	0	0	0
10	Hogg, 2000	0	0	0	0	0	0	0	0	46	6	46	7
11	Howell, 2001	0	0	0	0	0	0	0	0	169	16	171	13
12	Jain, 2003 *	72	11	0	0	0	2	36	7	0	0	0	0
13	Long, 2003	0	0	0	0	0	0	0	0	44	6	29	1
14	Loughnan, 2000	0	0	0	0	0	0	0	0	270	40	268	36
15	Lucas, 2001	0	0	0	0	0	0	0	0	304	62	309	63
16	Muir, 1996	0	0	0	0	0	0	0	0	20	2	25	3
17	Muir, 2000	0	0	0	0	0	0	0	0	79	9	86	11
18	Nafisi, 2006 *	179	19	0	0	0	0	173	24	0	0	0	0
19	Nikkola, 1997 *	6	0	4	0	0	0	10	0	0	0	0	0
20	Philipsen, 1989	0	0	0	0	0	0	0	0	48	6	47	10
21	Ramin, 1995 *	546	17	95	8	230	2	393	39	0	0	0	0
22	Sharma, 1997 *	336	16	5	0	114	1	231	12	0	0	0	0
23	Sharma, 2002	0	0	0	0	11	1	199	15	213	20	0	0
24	Shifman, 2007	0	0	0	0	0	0	0	0	32	18	45	15
25	Thalme, 1974	0	0	0	0	0	0	0	0	10	4	8	6
26	Thorp, 1993	0	0	0	0	0	0	0	0	44	1	36	12
27	Volmanen, 2008 *	23	1	3	0	1	0	23	1	0	0	0	0

The * indicates that the corresponding study has complete data on compliance status.

These RCTs varied in their inclusion criteria, labor management strategies, *etc.* In the 27 RCTs, 4,459 women were assigned to receive epidural analgesia and 4,426 were assigned to receive non-epidural or no analgesia. Complete data were available on the cesarean outcome, with 470 cesarean deliveries in women assigned to epidural and 419 cesarean deliveries in women assigned to non-epidural or no analgesia.

However, complete data on the number of cesarean sections in the compliant and non-compliant groups were available only for 10 studies, and data on noncompliance status per randomization group were only partly available for 13 of the 27 RCTs. We use $t = *$ to denote when the actually-received intervention is missing, then reorganize the available complete data and marginal data in Table 3.1.1. If N_{irto} were available for each $t \in \{0, 1\}$, the corresponding marginal count N_{ir*o} was assigned as 0; otherwise if the actual received intervention data for arm r of study i were missing, only the marginal data N_{ir*o} are shown in the table.

3.1.2 Analysis of Event Rates and Noncompliance Rates

Bannister-Tyrrell et al. (2015) estimated the association between epidural analgesia in labor and cesarean section using ITT and IV analysis, including only the 9 studies with complete data on the number of cesarean sections in compliant and noncompliant participants. We want to test whether the studies with incomplete data provide extra information. The ITT meta-analysis of the 27 RCTs gave a pooled Risk Ratio 1.10 (95% confidence interval: 0.97, 1.25; P=0.071) for cesarean section following epidural analgesia in labor, which implies that epidural analgesia in labor is not associated with risk of cesarean section. However, due to the high rate of noncompliance, an ITT analysis may not be a good way to estimate the effect of receiving epidural analgesia. The ITT pooled effect is potentially biased, especially when noncompliance cannot be assumed to be random with respect to the outcome in the meta-analysis. Therefore, we investigated the association between the ITT event rates and the noncompliance rates. Because several studies had 0 events or 0 noncompliance, we used a bivariate generalized linear mixed effects model (BGLMM) (Chu and Cole, 2006) to do the analysis on the 27 studies. The BGLMM assumes a bivariate normal distribution of probabilities in

the two groups (p_{1i}, p_{0i}) in a transformed scale, where the probabilities can be either event rates $\{p_{1i} = P(o_i = 1|r_i = 1),\ p_{0i} = P(o_i = 1|r_i = 0)\}$ or noncompliance rates $\{p_{1i} = P(t_i = 0|r_i = 1),\ p_{0i} = P(t_i = 1|r_i = 0),\}$, respectively. Specifically, the model is a probit model with:

$$\Phi^{-1}(p_{1i}) = u + \eta_{1i}, \quad \Phi^{-1}(p_{0i}) = v + \eta_{0i}, \quad (\eta_{1i}, \eta_{0i})^T \sim MVN(\mathbf{0}, \mathbf{\Sigma_\eta}). \tag{3.1}$$

In this model, $\Phi(.)$ is the standard Gaussian density function, (η_{1i}, η_{0i}) are random effects, and the covariance matrix is $\mathbf{\Sigma_\eta} = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}$. We chose the probit link because it has a closed-form formula for the overall probabilities $E(p_{1i}) = \Phi(u/\sqrt{1+\sigma_u^2})$ and $E(p_{0i}) = \Phi(v/\sqrt{1+\sigma_v^2})$, based on Equation (3.1).

We did a Bayesian analysis using JAGS (Plummer, 2003) to draw Markov chain Monte Carlo (MCMC) samples from the joint posterior distribution. We assigned vague priors $N(0, 1000)$ to the fixed effects u, v, and the commonly-used inverse Wishart distribution $InvW(I, \nu = 3)$ to the covariance matrix $\mathbf{\Sigma_\eta}$, where I is the identity matrix. The cesarean section event rates $\{p_{1i} = P(o_i = 1|r_i = 1),\ p_{0i} = P(o_i = 1|r_i = 0)\}$, and the intervention noncompliance rates $\{p_{1i} = P(t_i = 0|r_i = 1),\ p_{0i} = P(t_i = 1|r_i = 0)\}$ were analyzed separately using the model in Equation (3.1). After 10,000 burn-in samples, 40,000 posterior samples were obtained. The overall estimates $E(p_{1i})$, $E(p_{0i})$ were calculated using the closed form formula shown above. We write posterior medians followed by 95% equal-tail credible interval (CI) in brackets for the rest of this chapter. The overall probability of having a cesarean section in patients assigned to epidural analgesia was estimated as 12.9% (9.9%, 17.0%), while in those assigned to no/other analgesia it was 11.3% (8.5%, 15.0%). Also, the noncompliance rate in the epidural analgesia arm $E\{P(t_i = 0|r_i = 1)\}$ was 15.6% (5.4%, 29.0%), while while in the no/other analgesia arm $E\{P(t_i = 1|r_i = 0)\}$ was 13.8% (3.4%, 31.3%).

Figure 3.1 plots the study-specific posterior medians and 95% CIs for the cesarean section event rates (horizontal lines) and the noncompliance rates (vertical lines) in both the epidural analgesia arm (dashed line) and the control arm (solid line). Noncompliance rates show somewhat different patterns in the two randomization groups: as the event rate increases, the noncompliance rate tends to be higher, but this trend is more obvious in the control groups. Arguably, the association is in the opposite direction for the

treated groups.

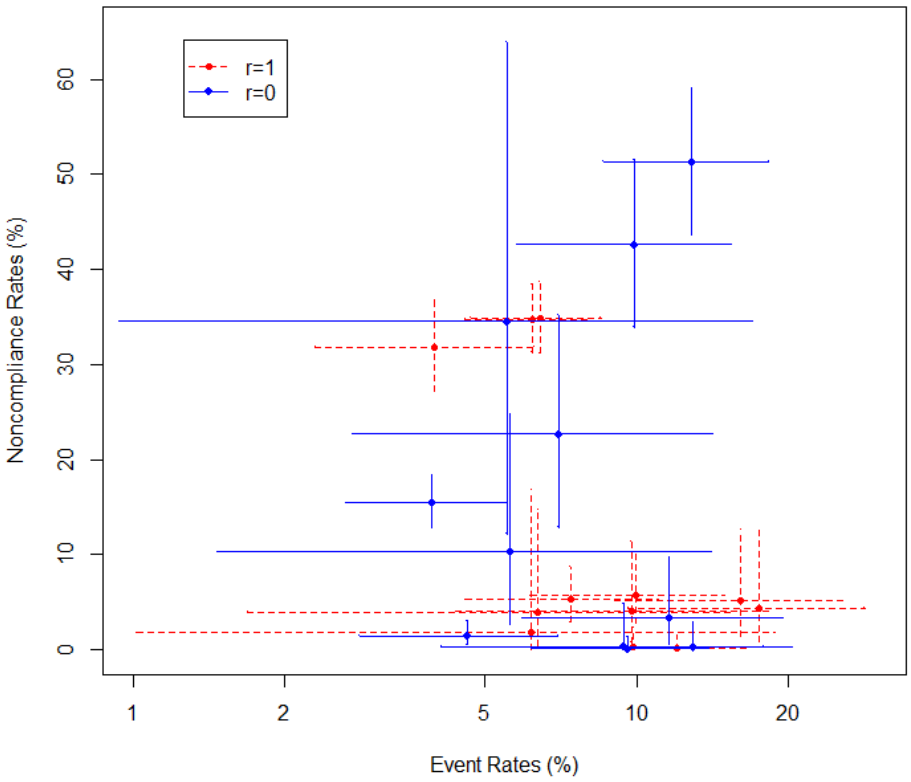

Figure 3.1: Study-specific event rates vs. noncompliance rates in studies of epidural analgesia in labor. Coordinates of each dot are the posterior medians of the study-specific event rate and compliance rate. Horizontal lines represent the 95% CI of study-specific posterior cesarean section event rate. Vertical lines represent the 95% CI of study-specific posterior noncompliance rate. Dashed lines show results in the epidural analgesia arm, while solid lines mark show results in no/other analgesia group. The horizontal axis has a logarithmic scale.

Therefore, the relationship between these two rates motivates us to do a causal

inference analysis on the treatment effects, rather than use the ITT analysis ignoring noncompliance. However, the existing complier average causal effect (CACE) framework needs complete information on compliance. With completely or partially missing data on compliance in many studies, we need a new method to include these studies but still give a valid causal interpretation. We introduce this method in Section 3.2 by first defining essential notations and assumptions.

3.2 Statistical Methods

3.2.1 Definition of the Complier Average Causal Effect (CACE)

Notation

In a meta-analysis with I two-armed randomized trials, N_i is the number of subjects in the i-th trial, where N_{i0} represents those who are randomly assigned to the control/placebo group and N_{i1} to the active treatment group. Let $R_{ij} = r$ index the randomization assignment for subject j in study i with $r = 0$ for assignment to control and $r = 1$ for assignment to treatment. Let $T_{ij}^{r} = t \in \{0, 1\}$ be the *potential* treatment received under the randomization assignment r, where $t = 1$ indicates receiving the active treatment and $t = 0$ placebo. Let $Y_{ij}^{r,t} = o \in \{1, 2, \ldots, O\}$ be the potential outcomes under randomization assignment r and treatment received t for the j-th subject in the i-th trial. Note that the sets of $\{Y_{ij}^{r,t}\}$ and $\{T_{ij}^{r}\}$ are the *potential* outcome and treatment-received status under possible r and t, but for each subject in a trial, only one of the possible values of each set can be observed. Therefore, we denote the observed response and received treatment variables as Y_{ij} and T_{ij} for the j-th subject in the i-th trial. We allow $T_{ij} = *$ if the actual received treatment is not recorded, and $Y_{ij} = *$ if the outcome is not recorded for the j-th patient in the i-th study. Then we let $\boldsymbol{M_i}$ be the N_i-dimensional vector of missing indicators for all subjects in trial i, with individual element $M_{ij} = m$ corresponding to whether subject j has actual treatment received status on record ($m = 0$) or missing ($m = 1$).

Following Imbens and Rubin (1997), we let C_{ij} be the latent compliance class of the

j-th patient in the i-th trial, defined as follows:

1) C_{ij}= 0, never-taker, if $(T_{ij}^0, T_{ij}^1) = (0, 0)$, *i.e.*, subjects who would receive control if randomized to either group;

2) C_{ij}= 1, complier, if $(T_{ij}^0, T_{ij}^1) = (0, 1)$, *i.e.*, subjects who would receive the intervention to which they were randomized;

3) C_{ij}= 2, always-taker, if $(T_{ij}^0, T_{ij}^1) = (1, 1)$, *i.e.*, subjects who would receive active treatment if randomized to either group;

4) C_{ij}= 3, defier, if $(T_{ij}^0, T_{ij}^1) = (1, 0)$, *i.e.*, subjects who would receive the intervention opposite to their randomized assignment.

A subject's compliance status C_{ij} is not observable because in a two-arm trial, only one of T_{ij}^1 and T_{ij}^0 can be observed. Based on the observed randomization group and actual treatment received, the compliance classes can only be partially identified (see Table 3.2.1, columns R_{ij}, T_{ij}, and C_{ij}).

Assumptions and Outcome Distributions

For each study, we make assumptions identical to those listed in Angrist et al. (1996):

Assumption 1: Stable unit treatment value assumption (SUTVA) (Rubin, 1980). The outcome for a subject is unaffected by the particular assignments of treatments to the other subjects. That is, if $r = r'$ then $T_{ij}^r = T_{ij}^{r'}$; and if $r = r'$ and $t = t'$ then $Y_{ij}^{r,t} = Y_{ij}^{r',t'}$.

Assumption 2: Random assignment to randomization groups. For all N_i subjects in the i-th trial, the treatment assignment is random. This assumption implies that the proportion of compliers should be the same in the intervention and control groups.

Assumption 3: Exclusion restriction. For subject j in the i-th trial, $Y_{ij}^{r,t} = Y_{ij}^{r',t}$ for all r, r' and t, *i.e.*, the randomization assignment affects responses only through its effect on treatment received. This assumption allows us to define $Y_{ij}^t \equiv Y_{ij}^{r,t} \equiv Y_{ij}^{r',t}$ for all r, r' and t. Therefore, for always-takers and never-takers, the distribution of outcomes does not depend on the randomization group.

Assumption 4: $E[T_{ij}^1 - T_{ij}^0] \neq 0$ *for each* i. For each trial, we assume the fraction of subjects who receive each intervention varies by randomization group.

Assumption 5: Monotonicity. $P[T_{ij}^1 \geq T_{ij}^0] = 1$ for each trial. This implies that no subject necessarily receives the treatment opposite to the assignment, under assignment to both active treatment and control. This assumption rules out the existence of defiers, and reduces the number of compliance types for which we must derive estimates, permitting a properly identified model.

Assuming randomized assignment and the exclusion restriction implies two restrictions: 1) the proportions of always-takers, never-takers and compliers in the control group are equal to those in the treatment group; 2) for never-takers and always-takers, the outcome distribution is the same under assignment to control and to active treatment. With these two restrictions, for discrete outcomes $o \in \{1, \ldots, O\}$ we can extend the notation in Cheng (2009) and Baker (2011), and define the following parameters for latent compliance classes and response rates in the i-th study: 1) π_{ia} and π_{in} are the probabilities of being an always-taker and a never-taker, respectively, so the probability of being a complier in the i-th study π_{ic} is $1 - \pi_{ia} - \pi_{in}$; 2) u_{io} is the probability of having outcome o for a complier randomized to the treatment group, and v_{io} is the probability for a complier randomized to the control/placebo group in the i-th study; s_{io} is the probability a never-taker has outcome o in the i-th study; and b_{io} is the probability an always-taker has outcome o in the i-th study; where $\sum_{o=1}^{O} u_{io} = \sum_{o=1}^{O} v_{io} = \sum_{o=1}^{O} s_{io} = \sum_{o=1}^{O} b_{io} = 1$. Although latent compliance classes cannot be fully identified based on randomization group (R_{ij}) and observed treatment received (T_{ij}), the above two restrictions allow us to write the distributions of observed N_{irt} in terms of the parameters for compliance classes and response rates, where $N_{irt} = \sum_j I(R_{ij} = r, T_{ij} = t)$ denotes the number of individuals in each observed group. Let $M(N_{irt}, \boldsymbol{x_{io}})$ denote a multinomial distribution with N_{irt} subjects and multinomial probabilities $\{\boldsymbol{x_{io}}\}$. The observed count for each outcome o in group $\{j : R_{ij} = r, T_{ij} = t\}$ is N_{irto}, $o = 1, \ldots, O$. Table 3.2.1 shows the distribution of each observed count in trial i, where $q_{io} = \frac{\pi_{ic} v_{io} + \pi_{in} s_{io}}{1 - \pi_{ia}}$ and $p_{io} = \frac{\pi_{ic} u_{io} + \pi_{ia} b_{io}}{1 - \pi_{in}}$ are probabilities corresponding to N_{i00o} and N_{i11o}, $o \in \{1, \ldots, O\}$.

Table 3.2: Observed groups, latent compliance classes and outcome probabilities of trial i

R_{ij}	T_{ij}	C_{ij}	$Y_{ij} = o \in \{1, \ldots, O\}$	Count
0	0	0 (never-taker) or 1 (complier)	$M(N_{i00}, q_{io} = \frac{\pi_{ic}v_{io}+\pi_{in}s_{io}}{1-\pi_{ia}})$	N_{i00o}
0	1	2 (always-taker) ~~or 3 (defier)~~	$M(N_{i01}, b_{io})$	N_{i01o}
1	0	0 (never-taker) ~~or 3 (defier)~~	$M(N_{i10}, s_{io})$	N_{i10o}
1	1	1 (complier) or 2 (always-taker)	$M(N_{i11}, p_{io} = \frac{\pi_{ic}u_{io}+\pi_{ia}b_{io}}{1-\pi_{in}})$	N_{i11o}

Defiers are ruled out by the monotonicity assumption.

Furthermore, according to the relations between observed groups and latent compliance classes, we have $\sum_o N_{i00o} = N_{i00} = N_{i0}(1-\pi_{ia})$ and $\sum_o N_{i01o} = N_{i01} = N_{i0}\pi_{ia}$, so the vector of observed counts in the control group $(N_{i001}, \ldots, N_{i00O}, N_{i011}, \ldots, N_{i01O})$ follows a multinomial distribution $M(N_{i0}, \boldsymbol{x_{i0}} = (x_{i001}, \ldots, x_{i00O}, x_{i011}, \ldots, x_{i01O}))$, where $x_{i00o} = q_{io}(1-\pi_{ia}) = \pi_{ic}v_{io}+\pi_{in}s_{io}$, $x_{i01o} = b_{io}\pi_{ia}$, and $o \in \{1, \ldots, O\}$. Similarly, in the active treatment group, the vector of observed counts $(N_{i101}, \ldots, N_{i10O}, N_{i111}, \ldots, N_{i11O})$ follows a multinomial distribution $M(N_{i1}, \boldsymbol{x_{i1}} = (x_{i101}, \ldots, x_{i10O}, x_{i111}, \ldots, x_{i11O}))$, where $x_{i10o} = s_{io}\pi_{in}$, $x_{i11o} = p_{io}(1-\pi_{ia}) = \pi_{ic}u_{io} + \pi_{ia}b_{io}$, and $o \in \{1, \ldots, O\}$.

Let λ_i be the probability $P(R_{ij} = 1)$, which is usually known in a trial and treated as fixed. Therefore, for study i $(i = 1, 2, \ldots, I)$, all observed counts N_{irto} follow a single multinomial distribution, with corresponding probability P_{irto}, for $r \in \{0, 1\}, t \in \{0, 1\}$, $o \in \{1, \ldots, O\}$. In mathematical notation, the distribution is $M(N_i, \boldsymbol{x_i} = \{P_{irto}\})$, where $P_{i0to} = (1-\lambda_i)x_{i0to}$ and $P_{i1to} = \lambda_i x_{i0to}$.

In addition to *Assumptions 1-5*, we make the latent ignorable (LI) missing assumption as described in Section 1.3. That is, given the observed data and the latent unobserved compliance classes, missingness has no residual dependence on the outcomes. Under the LI assumption, Table 3.2.1 summarizes a typical data structure and notation for a study i with missing treatment-received status for randomized treatment group $r \in \{0, 1\}$. In each cell of Table 3.2.1, the first row shows the count and the second row shows the corresponding probability of the outcome; for a study in which subjects

randomized to r had missing data on actual treatment received, only the rows labeled "Missing" would be observed.

Table 3.3: Typical data for study i with missing actual treatment received status in randomization group $r \in \{0, 1\}$

Treatment received	Outcome		
	1	...	O
0	N_{ir01}	...	N_{ir0O}
	P_{ir01}	...	P_{ir0O}
1	N_{ir11}	...	N_{ir1O}
	P_{ir11}	...	P_{ir1O}
Missing	N_{ir*1}	...	N_{ir*O}
	$P_{ir01} + P_{ir11}$	...	$P_{ir0O} + P_{ir1O}$

In each cell, the first row: the observed count;
second row: the corresponding probability.

CACE in Meta-analysis

One causal effect of interest in many studies is the CACE discussed in Section 1.3. CACE for the i-th two-arm trial is defined as $\theta_i^{\text{CACE}} = E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1)$. The overall causal effect θ^{CACE} from the meta-analysis can be estimated by taking the expectation of θ_i^{CACE} over all I trials, $\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}})$. For an ordinal outcome $Y_{ij} = o \in \{1, \ldots, O\}$, suppose we use equally spaced scores $\{1, 2, ..., O\}$ to reflect the real distances between categories, then θ_i^{CACE} is $\sum_o (o \times u_{io}) - \sum_o (o \times v_{io})$. When the outcome is binary, we let $o \in \{0, 1\}$, so the CACE for the i-th trial is $\theta_i^{\text{CACE}} = u_{i1} - v_{i1}$.

A positive (negative) value of θ_i^{CACE} indicates a beneficial treatment effect in the i-th trial if a higher value of o means a better (worse) outcome, and $\theta_i^{\text{CACE}} = 0$ indicates no causal effect of treatment for compliers. Besides the aforementioned equally spaced scores $\{1, 2, ..., O\}$, their linear transforms may also be sensible in many cases and provide a reasonable compromise (Agresti, 2003). Alternative scoring systems such

as midranks are also possible. When uncertain about which scoring choice to use, a sensitivity analysis can be conducted on different reasonable choices to see how they affect the estimates.

3.2.2 Estimation and Inference

The Likelihood

Let $\boldsymbol{N_i} = \{\boldsymbol{N_{ir}}\}$ be the vector of observed data in study i, where r refers to the randomization group ($r = 1$ for treatment and $r = 0$ for the control/placebo arm). In each arm r, $\boldsymbol{N_{ir}} = \{\boldsymbol{N_{ir}^c}, \boldsymbol{N_{ir}^m}\}$, where the superscripts c and m denote complete and marginal counts, respectively. $\boldsymbol{N_{ir}^c} = \{N_{irto}\}$ under each $t \in \{0, 1\}$, and $o \in \{1, \ldots, O\}$. If the full compliance data were observed in arm r of study i, the corresponding marginal counts $\boldsymbol{N_{ir}^m} = \{N_{ir*o}\}$ are assigned as 0. Otherwise, if the actual received-treatment status in randomization arm r of study i was missing, only the marginal data $\boldsymbol{N_{ir}^m} = \{N_{ir*o}\}$ are available.

From Section 3.1.1, if full compliance data were observed in both randomization groups, all observed counts N_{irto} follow a single multinomial distribution, with probability P_{irto}, where $P_{i0to} = (1-\lambda_i)x_{i0to}$ and $P_{i1to} = \lambda_i x_{i0to}$. Furthermore, as indicated by Table 3.2.1, all N_{ir*o} also follow a multinomial distribution with probability $P_{ir0o} + P_{ir1o}$ if only marginal data were observed, for $o \in \{1, \ldots, O\}$ in the i-th trial. Therefore, defining $\boldsymbol{\beta_i} = (\pi_{ia}, \pi_{in}, \boldsymbol{s_i}, \boldsymbol{b_i}, \boldsymbol{u_i}, \boldsymbol{v_i})$, where $\boldsymbol{s_i} = (s_{i1}, \ldots, s_{i(O-1)})$, $\boldsymbol{b_i} = (b_{i1}, \ldots, b_{i(O-1)})$, $\boldsymbol{u_i} = (u_{i1}, \ldots, u_{i(O-1)})$, $\boldsymbol{v_i} = (v_{i1}, \ldots, v_{i(O-1)})$, study i's likelihood contribution is

$$L_i(\boldsymbol{\beta_i}) = \prod_j \prod_o P_{i00o}^{(1-R_{ij})(1-T_{ij})(1-M_{ij})I(Y_{ij}=o)} P_{i01o}^{(1-R_{ij})T_{ij}(1-M_{ij})I(Y_{ij}=o)} P_{i10o}^{R_{ij}(1-T_{ij})(1-M_{ij})I(Y_{ij}=o)} P_{i11o}^{R_{ij}T_{ij}(1-M_{ij})I(Y_{ij}=o)}$$

$$(P_{i00o} + P_{i01o})^{(1-R_{ij})M_{ij}I(Y_{ij}=o)} (P_{i10o} + P_{i11o})^{R_{ij}M_{ij}I(Y_{ij}=o)}, \tag{3.2}$$

where the relations among the components of $\boldsymbol{\beta_i}$ and P_{irto} are summarized in Section 3.2.1, $j = 1, \ldots, N_i$, $o = 1, \ldots, O$, and the indicator function $I(Y_{ij} = o) = 1$ if $Y_{ij} = o$ and 0 otherwise. The parameters are subject to $\sum_o u_{io} = \sum_o v_{io} = \sum_o s_{io} = \sum_o b_{io} = 1$

and $0 \leq \pi_{ia}, \pi_{in}, u_{io}, v_{io}, s_{io}, b_{io} \leq 1$. The likelihood function for all trials in a meta-analysis is $\mathcal{L}(\boldsymbol{\beta}) = \prod_i L_i(\boldsymbol{\beta_i})$.

We use trials with binary outcomes to further illustrate the modeling; this also represents the situation in the motivating example. In this case, $o \in \{0, 1\}$, *i.e.*, $s_{i0} + s_{i1} = b_{i0} + b_{i1} = u_{i0} + u_{i1} = v_{i0} + v_{i1} = 1$ for study i, so the vector parameters of $\boldsymbol{s_i}, \boldsymbol{b_i}, \boldsymbol{u_i}, \boldsymbol{v_i}$ are reduced to $s_{i1}, b_{i1}, u_{i1}, v_{i1}$. Data can be arranged as shown in Table 3.1.1, where in each randomization arm, data are shown either in the column "Complete data" or in the column "Missing data", with values in the other columns all 0. Thus the observed data are $\boldsymbol{N_{ir}} = \{\boldsymbol{N_{ir}^c}, \boldsymbol{N_{ir}^m}\} = \{N_{ir00}, N_{ir01}, N_{ir10}, N_{ir11}, N_{ir*0}, N_{ir*1}\}$ for $r \in \{0, 1\}$. Then the likelihood contribution for the i-th trial can be written as

$$
\begin{aligned}
L_i(\boldsymbol{\beta_i}) = {} & [(1-\lambda_i)\{\pi_{ic}(1-v_{i1}) + \pi_{in}(1-s_{i1})\}]^{N_{i000}} \{(1-\lambda_i)(\pi_{ic}v_{i1} + \pi_{in}s_{i1})\}^{N_{i001}} \\
& \{(1-\lambda_i)\pi_{ia}(1-b_{i1})\}^{N_{i010}} \{(1-\lambda_i)\pi_{ia}b_{i1}\}^{N_{i011}} \{\lambda_i\pi_{in}(1-s_{i1})\}^{N_{i100}} \\
& \{\lambda_i\pi_{in}s_{i1}\}^{N_{i101}} [\lambda_i\{(\pi_{ic}(1-u_{i1}) + \pi_{ia}(1-b_{i1})\}]^{N_{i110}} \{\lambda_i(\pi_{ic}u_{i1} + \pi_{ia}b_{i1})\}^{N_{i111}} \\
& [(1-\lambda_i)\{\pi_{ic}(1-v_{i1}) + \pi_{in}(1-s_{i1}) + \pi_{ia}(1-b_{i1})\}]^{N_{i0*0}} \{(1-\lambda_i)(\pi_{ic}v_{i1} + \pi_{in}s_{i1} + \pi_{ia}b_{i1})\}^{N_{i0*1}} \\
& [\lambda_i\{\pi_{ic}(1-u_{i1}) + \pi_{ia}(1-b_{i1}) + \pi_{in}(1-s_{i1})\}]^{N_{i1*0}} \{\lambda_i(\pi_{ic}u_{i1} + \pi_{ia}b_{i1} + \pi_{in}s_{i1})\}^{N_{i1*1}}
\end{aligned}
\tag{3.3}
$$

where $\boldsymbol{\beta_i} = (\pi_{ia}, \pi_{in}, s_{i1}, b_{i1}, u_{i1}, v_{i1})$, and the parameters vary between studies following some distributions with hyper-parameters, which we now describe.

To account for potential between-study heterogeneity of the compliance classes and outcome probabilities, we consider a random effects model. Specifically, to guarantee the desired properties of latent compliance classes in study i, *i.e.*, $\pi_{in} + \pi_{ia} + \pi_{ic} = 1$ and $0 \leq \pi_{in}, \pi_{ia}, \pi_{ic} \leq 1$, and to allow these probabilities to vary between studies, the parameters are specified as: $\pi_{in} = \frac{exp(n_i)}{1+exp(n_i)+exp(a_i)}, \pi_{ia} = \frac{exp(a_i)}{1+exp(n_i)+exp(a_i)}$, where $n_i = \alpha_n + \delta_{in}, a_i = \alpha_a + \delta_{ia}$. The random effect $(\delta_{in}, \delta_{ia})$ has a bivariate normal distribution with mean 0 and variance-covariance matrix $\boldsymbol{\Sigma}_{lc} = \begin{pmatrix} \sigma_n^2 & \rho\sigma_n\sigma_a \\ \rho\sigma_n\sigma_a & \sigma_a^2 \end{pmatrix}$, to allow correlation between n_i and a_i.

We also define random effect models on the transformed scale of each response probability $s_{i1}, b_{i1}, u_{i1}, v_{i1}$: $g(s_{i1}) = \alpha_s + \delta_{is}$, $g(b_{i1}) = \alpha_b + \delta_{ib}$, $g(u_{i1}) = \alpha_u + \delta_{iu}$, $g(v_{i1}) =$

$\alpha_v + \delta_{iv}$, where $g(\cdot)$ is a link function such as the logit or probit. These response rates are assumed to be independent across principal strata, so $\delta_{is} \sim N(0, \sigma_s^2)$, $\delta_{ib} \sim N(0, \sigma_b^2)$, $\delta_{iu} \sim N(0, \sigma_u^2)$, $\delta_{iv} \sim N(0, \sigma_v^2)$. The model can easily be extended to more general cases with more than binary outcomes.

Prior Specifications and the Posterior Distribution

We assign proper but diffuse prior distributions for the hyper-parameters. Specifically, α_n and α_a both follow $N(0, 2.5^2)$, such that under the simplest situation (a fixed effects model), a 95% prior probability interval for any of the probabilities $\pi_{in}, \pi_{ia}, \pi_{ic}$ ranges from about 0.001 to 0.91; and $\alpha_s, \alpha_b, \alpha_u, \alpha_v$ all follow $N(0, 2^2)$, which implies a 95% interval for the probabilities $s_{i1}, b_{i1}, u_{i1}, v_{i1}$ ranging from about 0.01 to 0.98. The hyper-priors for the precision parameters σ_s^{-2}, σ_b^{-2}, σ_u^{-2} and σ_v^{-2} are assumed to be $Gamma(2, 2)$, which corresponds to a 95% interval of $(0.6, 2.9)$ for the corresponding standard deviations, allowing moderate heterogeneity in the response probabilities. The prior for the precision matrix $\mathbf{\Sigma}_{lc}^{-1}$ is Wishart, *i.e.*, $W(I, 3)$, where I is the identity matrix. In a reduced model with one of σ_n^2, σ_a^2 set to 0, the prior of the other precision parameter is also assumed to be $Gamma(2, 2)$, which gives moderate heterogeneity for latent compliance classes probabilities.

Let function $f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \mathbf{\Sigma_0})$ be the distributions described in Section 3.2.2 of all parameters $\boldsymbol{\beta_i} = (\pi_{ia},\ \pi_{in},\ s_{i1},\ b_{i1},\ u_{i1},\ v_{i1})$, where $\boldsymbol{\beta_0}$ refers to the vector of mean hyper-parameters $(\alpha_n,\ \alpha_a,\ \alpha_s,\ \alpha_b,\ \alpha_u,\ \alpha_v)$, and $\mathbf{\Sigma_0}$ is the covariance matrix of hyper-parameters $\mathbf{\Sigma}_{lc}^{-1}$, σ_s^{-2}, σ_b^{-2}, σ_u^{-2} and σ_v^{-2}. Denoting the prior distributions specified above as $f(\boldsymbol{\beta_0})$ and $f(\mathbf{\Sigma_0})$, the joint posterior distribution is then proportional to $\prod_i L_i(\boldsymbol{\beta_i}) f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \mathbf{\Sigma_0}) f(\boldsymbol{\beta_0}) f(\mathbf{\Sigma_0})$. We sample from the joint posterior using Markov chain Monte Carlo (MCMC) methods, specifically Gibbs and Metropolis-Hastings sampling algorithms (Gelfand and Smith, 1990).

As mentioned in Section 3.2.1, for binary outcomes, θ^{CACE} can be estimated as $E(\theta_i^{\text{CACE}}) = E(u_{i1}) - E(v_{i1})$. Integrating out the random effects, $E(u_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t)\sigma_u^{-1}$ $\phi(\frac{t}{\sigma_u})dt$ and $E(v_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_v + t)\sigma_v^{-1}\phi(\frac{t}{\sigma_v})dt$, where $\phi(\cdot)$ is the standard Gaussian density. Using probit link functions for u_{i1} and v_{i1}, we have closed-form formulas

$E(u_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}})$ and $E(v_{i1}) = \Phi(\frac{\alpha_v}{\sqrt{1+\sigma_v^2}})$ so that

$$\theta^{\text{CACE}} = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}}) - \Phi(\frac{\alpha_v}{\sqrt{1+\sigma_v^2}}) \tag{3.4}$$

For s_{i1} and b_{i1}, we used the logit link random effects model. Though the integral $E(s_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_s + t)\sigma_s^{-1}\phi(\frac{t}{\sigma_s})dt$ does not have a closed-form formula, it has a well-established approximation, $E(s_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_s}{\sqrt{1+C^2\sigma_s^2}})$, where $C = \frac{16\sqrt{3}}{15\pi}$ (Zeger et al., 1988). This approximation also applies to estimating the overall always-taker response rate $E(b_{i1})$.

In each MCMC iteration, draws of θ^{CACE} are calculated from the MCMC draws using Equation (3.4). We use medians and equal-tail credible intervals (CIs) of these posterior samples to make inferences for the random effects models.

Model Selection and Implementation

The model specified in Section 3.2.2 included all possible random effects to account for possible between-study heterogeneity of the fractions in the compliance classes and of the response rate probabilities. However, over-fitting the data with too many random effects should be avoided because it may inflate posterior variances. Therefore, we have used a forward selection procedure to choose the final model, beginning with a model having no random effects and at each forward step adding the random-effect component that gave the largest improvement in the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Other model-selection approaches can be substituted easily, *e.g.*, using a different model-selection criterion or a different search strategy.

We used JAGS software version 4.3 *via* the rjags package in R to sample from the joint posterior distribution. We ran three independent MCMC chains with starting points drawn randomly from their prior distributions. After 10,000 burn-in samples, the subsequent 100,000 posterior samples were obtained for each chain. Convergence to the stationary distribution was assessed using trace plots, sample autocorrelation, and the Gelman and Rubin statistic (Gelman and Rubin, 1992).

Model for Complete Data Only

Here we discuss how the naive "two-step" approach introduced in Section 1.3 can be viewed as a special case of our model using only trials with complete noncompliance data. In this situation, only trials with complete data $\boldsymbol{N_{ir}^c} = \{N_{irto}\}$ are used to make inference on CACE. Then the likelihood for the i-th study is

$$L_i(\boldsymbol{\beta_i}) = \prod_j \prod_o P_{i00o}^{(1-R_{ij})(1-T_{ij})(1-M_{ij})I(Y_{ij}=o)} P_{i01o}^{(1-R_{ij})T_{ij}(1-M_{ij})I(Y_{ij}=o)} P_{i10o}^{R_{ij}(1-T_{ij})(1-M_{ij})I(Y_{ij}=o)} P_{i11o}^{R_{ij}T_{ij}(1-M_{ij})I(Y_{ij}=o)}. \quad (3.5)$$

Note that when $M_{ij} = 1$ (i.e., for trials with incomplete noncompliance data), $L_i(\boldsymbol{\beta_i}) = 0$. Thus for trial i with complete noncompliance data $\boldsymbol{N_{ir}^c} = \{N_{irto}\}$, one can separately estimate θ_i^{CACE} and obtain a standard error. One can then combine these study-specific estimates using a standard meta-analysis method, such as a fixed-effect or random effects model, to estimate the population-averaged CACE. Alternatively, one can obtain the posterior estimate of θ^{CACE} through the joint posterior distribution, which is proportional to the likelihood for trials with complete noncompliance data $\mathcal{L}(\boldsymbol{\beta}) = \prod_i L_i(\boldsymbol{\beta_i})$ multiplied by the prior distributions. Note that by Lin and Zeng (2010), the two-step approach can be viewed as asymptotically equivalent to the model maximizing the joint likelihood. Therefore, in the simulation section below, we compare the performance of our proposed model including all trials with a model using only trials with complete noncompliance data instead of a two-step frequentist approach.

3.3 Case Study Results

3.3.1 Model Selection Results

We estimated the CACE of epidural analgesia in labor on cesarean section by including all of the 27 RCTs introduced in Section 3.1. Although the full model has 6 potential random effects in total, δ_{in}, δ_{ia}, δ_{is}, δ_{ib}, δ_{iu} and δ_{iv}, we adopted the forward selection procedure described in Section 3.2.2. Table 3.3.1 presents DIC, DIC improvement, and the effective number of parameters (p_D) for each model considered in the forward selection

procedure. Starting with the model that includes no random effects (called Model I), the largest improvement was obtained by adding a random effect for the study-specific probability of being an always-taker, *i.e.*, δ_{ia}, called Model IIf. In the next step, adding a random effect for the probability of being a never-taker, δ_{in} (Model IIIe), reduced DIC the most. This revealed an important characteristic of this meta-analysis: the studies might vary considerably in their recruitment criteria, study procedures, beliefs of the local PIs, *etc.*, so that properties of the latent compliance classes vary between studies. The next forward step produced a meaningful improvement by including a random effect for the cesarean section rate of a never-taker, s_{i1} (Model IVa). For the fourth forward step, DIC was improved most by adding one more random effect, the probability of having a cesarean section for a complier in the randomized treatment group, δ_{iu} (Model Vc). This improvement was modest compared with the previous step, but still notable (DIC decreased by 6.7 points compared with Model IVa). It is difficult to say what constitutes an important improvement in DIC; we follow Lunn et al. (2012) in considering that a reduction of less than 5 is not a substantial improvement. The final step found no notable improvements from including additional random effects. Therefore, the final model was Model Vc, including random effects δ_{ia}, δ_{in}, δ_{is}, and δ_{iu}.

Table 3.4: Selecting random effects using a forward selection procedure for the epidural analgesia in labor meta-analysis

Random effects models	DIC	DIC improvement	p_D
Model **I** (None)	**1409.0**	N/A	6.8
Model IIa (δ_{is})	1225.2	183.8	36.1
Model IIb (δ_{ib})	1258.5	150.5	33.2
Model IIc (δ_{iu})	1325.9	83.1	25.6
Model IId (δ_{iv})	1321.4	87.6	27.0
Model IIe (δ_{in})	992.5	416.5	89.6
Model **IIf** (δ_{ia})	**814.2**	**594.8**	26.2
Model IIIa (δ_{ia}, δ_{is})	748.3	65.9	45.0
Model IIIb (δ_{ia}, δ_{ib})	781.8	32.4	37.0
Model IIIc (δ_{ia}, δ_{iu})	790.9	23.3	37.6
Model IIId (δ_{ia}, δ_{iv})	804.6	9.6	39.2
Model **IIIe** (δ_{in}, δ_{ia})	**508.2**	**306.2**	33.3
Model **IVa** ($\delta_{in}, \delta_{ia}, \delta_{is}$)	**464.3**	**43.9**	46.2
Model IVb ($\delta_{in}, \delta_{ia}, \delta_{ib}$)	490.7	17.5	41.6
Model IVc ($\delta_{in}, \delta_{ia}, \delta_{iu}$)	494.3	13.9	47.7
Model IVd ($\delta_{in}, \delta_{ia}, \delta_{iv}$)	514.4	-6.2	49.1
Model IVe ($\delta_{in}, \delta_{ia}, \rho$)	507.8	0.4	34.0
Model Va ($\delta_{in}, \delta_{ia}, \delta_{is}, \rho$)	465.4	-1.1	46.9
Model Vb ($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{ib}$)	465.1	-0.7	49.9
Model **Vc** ($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$)	**457.7**	**6.7**	59.9
Model Vd ($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iv}$)	478.0	-13.7	60.3
Model VIa ($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}, \delta_{ib}$)	457.1	0.6	62.9
Model VIb ($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}, \delta_{iv}$)	465.9	-8.2	70.2

Bold-face cells are models selected using the forward selection procedure. DIC improvement is the reduction in DIC compared to the previous step's lowest DIC.

The study-specific component corresponding to each random effect: δ_{is} never-taker response, δ_{ib} always-taker response, δ_{iu} treated complier response, δ_{iv} control complier response, δ_{in} never-taker probability, δ_{ia} always-taker probability; ρ is the correlation between δ_{ia} and δ_{in}.

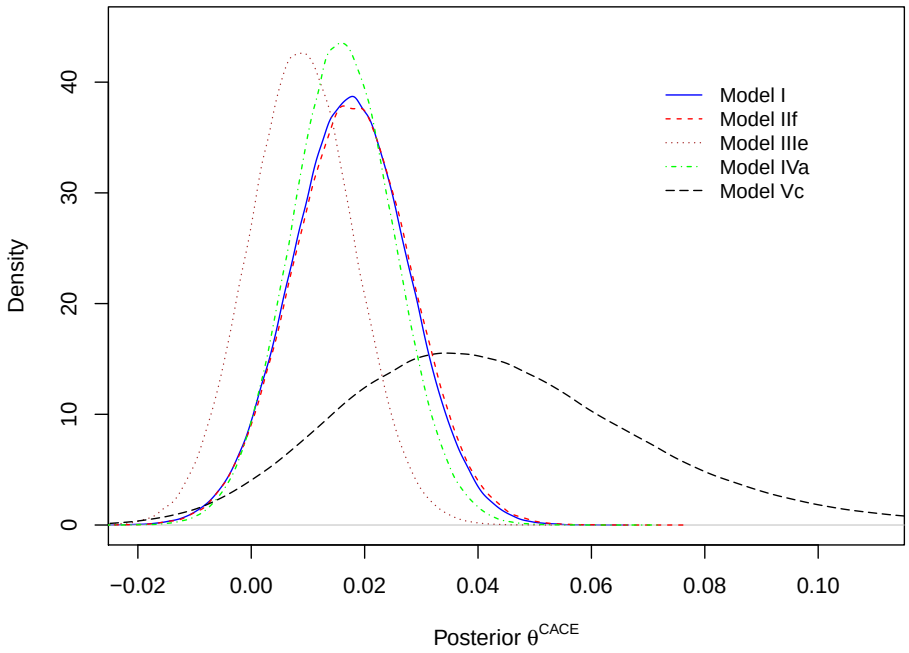

Figure 3.2: Posterior distributions of θ^{CACE} of epidural analgesia in labor on cesarean section, the kernel smoothed density estimate from 100,000 Monte Carlo samples.

Figure 3.2 shows the kernel-smoothed posterior density of θ^{CACE} from the Bayesian hierarchical models selected in each forward step. The plot suggests a fairly symmetric posterior density of θ^{CACE} for all models. After adding the random effect δ_{iu} to probit(u_{i1}) in Model Vc, the posterior of θ^{CACE} is shifted right and its variance increased considerably, which further indicates the importance of appropriate accounting for random effects.

Table 3.5: Summary of parameter estimates for the epidural analgesia meta-analysis

Parameter	Model I(None)	Model Vc($\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$)
θ^{CACE}	$_{-0.003}0.017_{0.038}$	$_{-0.003}0.041_{0.105}$
Overall never-taker probability π_n	$_{0.214}0.230_{0.246}$	$_{0.033}0.101_{0.259}$
Overall always-taker probability π_a	$_{0.136}0.152_{0.170}$	$_{0.065}0.190_{0.400}$
Overall complier probability π_c	$_{0.594}0.618_{0.641}$	$_{0.544}0.687_{0.787}$
Overall never-taker response s_1	$_{0.029}0.046_{0.068}$	$_{0.116}0.254_{0.488}$
Always-taker response b_1	$_{0.124}0.168_{0.216}$	$_{0.100}0.140_{0.174}$
Treated complier response u_1	$_{0.093}0.112_{0.131}$	$_{0.065}0.108_{0.173}$
Control complier response v_1	$_{0.078}0.095_{0.112}$	$_{0.054}0.068_{0.083}$
Mean parameter of n_i	$_{-1.089}$-0.988$_{-0.887}$	$_{-3.196}$-2.173$_{-1.224}$
Mean parameter of a_i	$_{-1.542}$-1.399$_{-1.260}$	$_{-3.521}$-2.038$_{-0.758}$
Standard deviation of n_i	–	$_{1.055}1.645_{2.846}$
Standard deviation of a_i	–	$_{1.402}2.240_{3.901}$
Standard deviation of logit(s_{i1})	–	$_{1.231}2.110_{4.131}$
Standard deviation of probit(u_{i1})	–	$_{0.431}0.600_{0.912}$

The notation ${}_L P_U$ denotes the posterior median P with 95% equal tailed credible limits (L, U).

Table 3.3.1 lists estimated parameters from the fixed effects model (Model I) and the final model (Model Vc), where the triple of percentiles, ${}_{2.5}50_{97.5}$, is used to display each parameter's posterior median with its 95% equal tail credible interval, as suggested by Louis and Zeger (2009). Monte Carlo integration (Ueberhuber, 1997) was used to estimate the overall probabilities of being in each principal stratum, π_a, π_c, and π_n when δ_{in} and δ_{ia} were both present (Model Vc). The overall never-taker response rate $s_1 = E(s_{i1})$ of Model Vc was estimated using the approximation $E(s_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_s}{\sqrt{1+C^2\sigma_s^2}})$, $C = \frac{16\sqrt{3}}{15\pi}$, and the overall treated complier response rate $u_1 = E(u_{i1})$ was estimated using the closed-form formula $E(u_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}})$. For other overall response rates (e.g., b_1, v_1), the values were directly estimated by transforming back the fixed-effect

parameters if the probabilities were assumed to be the same across studies according to either Model I or Model Vc. For example, the overall always-taker response rate was $b_1 = E(b_{i1}) = \text{logit}^{-1}(\alpha_b)$ because b_{i1} had no random effect in either model. Based on the final model (Model Vc), the posterior median and interval for θ^{CACE} were ${}_{-0.003}0.041_{0.105}$, which covers zero and indicates a nonsignificant complier average causal effect, but the estimated effect was about 2 that estimated by the fixed effect model (Model I). The random effects for π_a, π_n, and π_c on the transformed scale had standard deviations of 1.65 and 2.24, while the random effect for s_1 had a standard deviation of 2.11 on the logit scale. After adding random effects for δ_{in}, δ_{ia}, δ_{is} and δ_{iu}, the posteriors of π_n, π_a, s_1 and u_1 changed markedly from those estimated by Model I.

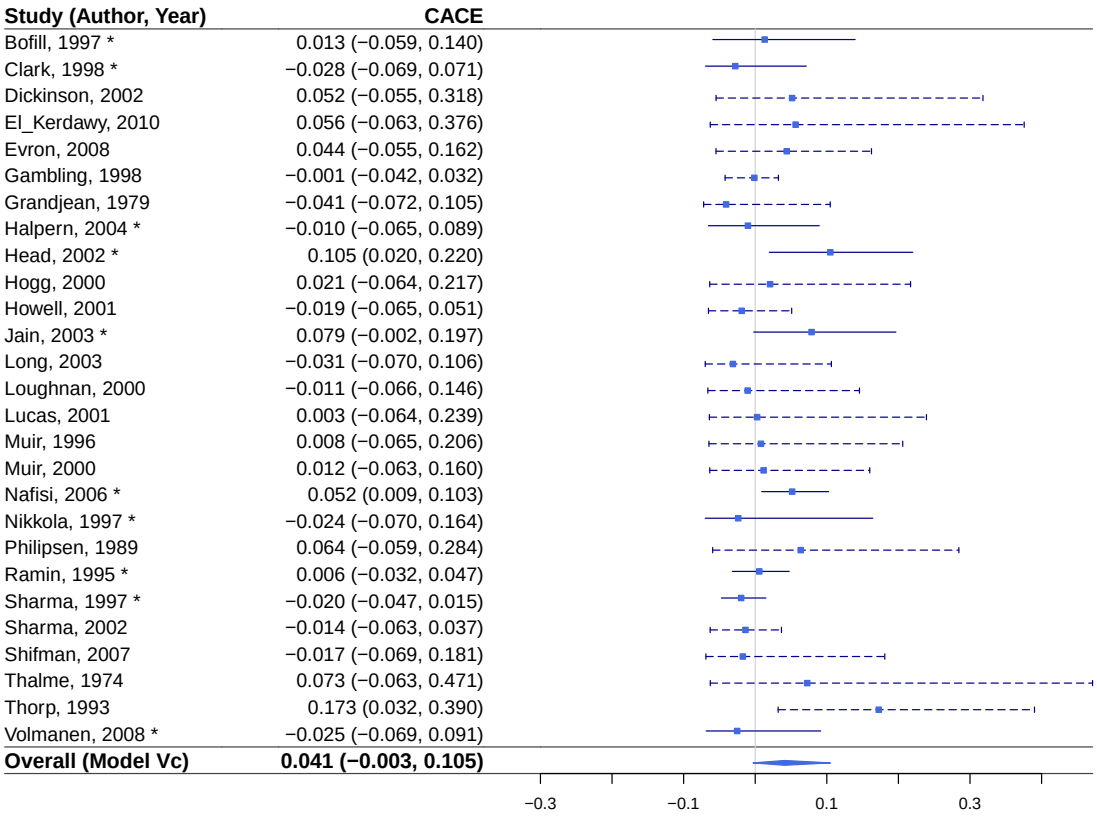

Figure 3.3: Forest plot of θ^{CACE} of epidural analgesia in labor on cesarean section. The center of each square and the horizontal lines represent the posterior median and 95% equal tail credible interval of θ_i^{CACE} for each study from the final model, Model Vc. The diamond indicates the pooled estimate of θ^{CACE} and its 95% credible interval. The symbol * indicates that the study has complete data on compliance status. With complete data, a solid horizontal line is used to represent the posterior 95% CI of θ_i^{CACE}, whereas a dashed line is used to show the CI for a study with incomplete compliance data.

Figure 3.3 is a forest plot of the posterior medians and 95% equal-tail credible intervals of θ_i^{CACE} for each trial based on the final model, Model Vc. Studies with a

"*" in the "Study (Author, Year)" column had complete data on compliance status and we used solid horizontal lines to represent their corresponding CIs. For a study with incomplete data, as its θ_i^{CACE} was not directly estimable by the single trial, we used a dashed line to show the posterior 95% CI. The figure showed that studies with complete data tend to have shorter credible intervals, while the study-specific estimates θ_i^{CACE} were quite heterogeneous, indicating differences in the study populations. The overall θ^{CACE} from the final model was not significant.

3.3.2 Sensitivity to the LI Assumption

The above models were built upon the assumption of latent ignorable (LI) missingness. However, this assumption may not be satisfied in some applications. For example, studies showing a treatment effect may have a higher chance of reporting compliance status. This is a form of missing not at random (MNAR): the probability of missing compliance data depends on the outcome measure. However, in practice, one can never tell from the data at hand whether missingness is LI or MNAR (Little and Rubin, 2014). Thus, we present a sensitivity analysis that incorporates a known MNAR mechanism to see its impact on treatment-effect estimates.

Let the $I \times 2$ matrix Ξ denote the study-level compliance missingness of a meta-analysis dataset containing I studies and 2 treatment arms. The entries of Ξ are ξ_{ir}, $i = 1, \ldots, I$ and $r = 0, 1$, such that $\xi_{ir} = 1$ if compliance information is missing in randomized group r of study i, and $\xi_{ir} = 0$ if the data is complete. We assume $\xi_{ir} \sim \text{Bern}(p_{ir}^{\text{mis}})$, where p_{ir}^{mis} is the probability of missing compliance status (*i.e.*, no data on the actual treatment taken) in study i's randomized group r. We specify a model of missingness for p_{ir}^{mis} as

$$\text{logit}(p_{i0}^{\text{mis}}) = \gamma_{00} + \gamma_{10} \times \text{logit}(v_{i1}),\ \text{logit}(p_{i1}^{\text{mis}}) = \gamma_{01} + \gamma_{11} \times \text{logit}(u_{i1}), \tag{3.6}$$

In this model, γ_{00} (γ_{01}) is an unknown scalar parameter, and γ_{10} (γ_{11}) describes the strength of association between the missingness probability and the study-specific response rate of a complier in the randomized control (treatment) group, *i.e.*, the components of θ_i^{CACE}. When $\gamma_{1r} = 0$ for $r = 0, 1$, the missingness probabilities are not related

to any model parameters, hence the missingness is completely at random (MCAR). For the purpose of assessing the effect of MNAR, for some given γ_{10} and γ_{11}, this model of missingness can be incorporated in the likelihood in Section 3.2.2 and treated as if it is known to be true. Note that the model of missingness described here is not for general MNAR scenarios, but is specific for the CACE problem, and we only consider scenarios in which missingness is related to components of θ^{CACE}.

In this case study, as the random effect δ_{iv} was not selected into the final Model Vc, the response rates for compliers randomized to the control group (v_{i1}) were the same across trials. Thus the missing probabilities in the control arm p_{i0}^{mis} according to Equation (3.6) were also the same for all studies i. For illustration, we set $\gamma_{10} = 0$ in conducting sensitivity analyses to the specific MNAR scenario, to explore the impact on CACE estimates as γ_{11} changes from negative to positive. Since a flat prior for γ_{0r} with a large variance would lead to a marginal prior distribution for p_{ir}^{mis} heavily weighted towards 0 and 1, we follow Zhang et al. (2017) by specifying a $logistic(0, 1)$ prior for γ_{0r}, which gives an approximate uniform prior for p_{ir}^{mis} on $(0, 1)$.

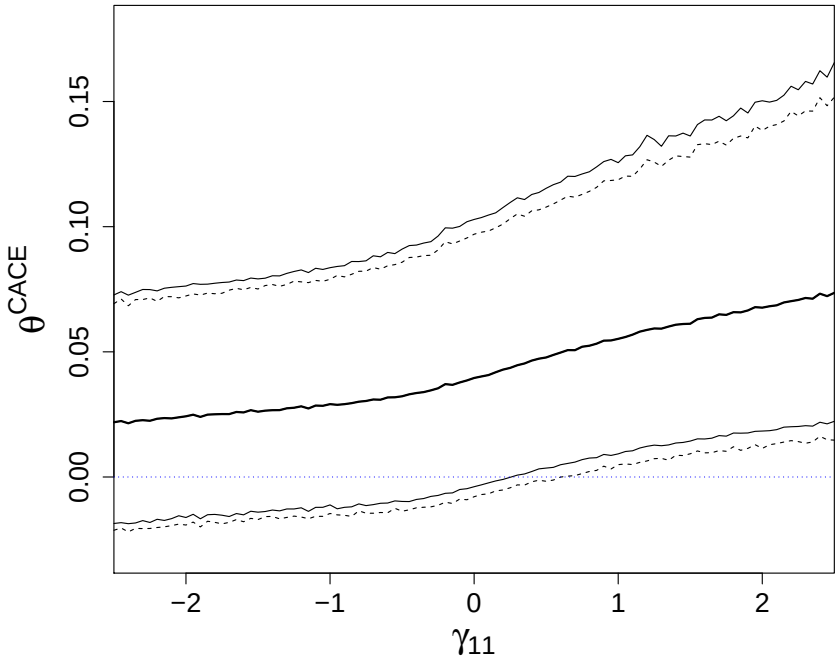

Figure 3.4: Posterior of θ^{CACE} of the epidural analgesia in labor meta-analysis under the assumption that the missingness probability in the treatment arm p_{i1}^{mis} is linearly related to u_{i1} on the logit scale. Bold solid line: posterior median; fine solid lines: 95% equal-tail credible interval; fine dashed lines: 95% highest posterior density credible interval. The fine dotted horizontal line is $\theta^{\text{CACE}} = 0$.

Figure 3.4 summarizes the posterior of θ^{CACE} from the meta-analysis of epidural analgesia in labor when we set $\gamma_{01} = 0$ in Equation (3.6) and allowed γ_{11} to range from -2.5 to 2.5 under the final model (Model Vc). As γ_{11} increased from -2.5 to 2.5, the posterior median of θ^{CACE} increased from about 0.02 to 0.07, and the 95% credible interval of θ^{CACE} no longer covered zero when the coefficient of $\text{logit}(u_{i1})$ was over about 0.5. That is to say, when the missingness probabilities were positively and strongly enough correlated with u_{i1}, the CACE became statistically significant, which differs from the conclusion drawn in Section 3.3.1 under the LI assumption. Therefore, the missingness mechanism for compliance would influence the causal effect estimates

in this epidural analgesia in labor meta-analysis.

3.4 Simulation

3.4.1 Simulation Setups

Simulation studies were conducted to evaluate how the proposed method performs under different assumptions. As in the case study, we assumed $o \in \{0, 1\}$, *i.e.*, the outcome is binary. We set $(\alpha_n, \alpha_a, \alpha_s, \alpha_b, \alpha_u, \alpha_v) = (-0.4, -0.6, 0.5, -0.5, -0.5, 0.5)$, so that the true values in the absence of random effects were $\pi_{ic} = 0.45$, $\pi_{in} = 0.30$, $\pi_{ia} = 0.25$ and $\theta_i^{\text{CACE}} = -0.38$. When random effects were present, we assumed the random effects had standard deviation 0.5, *i.e.*, each of $\sigma_n, \sigma_a, \sigma_s, \sigma_u = 0.5$. To evaluate the model's performance and the impact of random effects, we generated compliance status and outcomes data with three sets of random effects, corresponding to Section 3.3.1's Model IIIe $(\delta_{in}, \delta_{ia})$, Model IVa $(\delta_{in}, \delta_{ia}, \delta_{is})$, and Model Vc $(\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu})$. Under each scenario, we simulated 2000 datasets. Each dataset comprised 20 studies where 350 subjects per study were randomized to either the treatment or control group with a $1:1$ ratio ($\lambda = 0.5$).

We created partially missing compliance data under the MCAR, LI, and MNAR assumptions, as follows. Under the MCAR assumption, the missing indicators for all studies were prespecified such that the first ten studies in the control arm ($R = 0$), and the 6-th to the 15-th studies in the treatment arm ($R = 1$) did not have compliance information, so that only 5 studies had full data in both arms. To generate partially incomplete data under the LI and MNAR assumption, we applied a logit model to calculate the missingness probabilities in the control arm ($R = 0$) and treatment arm ($R = 1$) separately, which were used to generate the random missingness indicators to keep only the marginal data in that arm of the study. The models for the missingness indicators are:

$$\xi_{ir} \sim \text{Bern}(p_{ir}^{\text{mis}}),$$

$$\text{LI:}\ \ \text{logit}(p_{ir}^{\text{mis}}) = \beta_{0r} + \beta_{1r} \times \text{logit}(\pi_{ic})$$

$$\text{MNAR:}\ \ p_{i0}^{\text{mis}} = 0.5,\ \ \text{logit}(p_{i1}^{\text{mis}}) = \gamma_{01} + \gamma_{11} \times \text{logit}(u_{i1}),$$

where $r = 0, 1$ indicate the control and treatment groups respectively. If the missing indicator $\xi_{ir} = 1$, then data on compliance status in the i-th study arm r were set to be missing, *i.e.*, only marginal values N_{ir*1}, N_{ir*0} were available. Our model settings imply that the parameter π_{ic} is independent of θ_i^{CACE}, so we considered the missing assumption to be LI. For ease of presentation, we let $p_{i0}^{\text{mis}} = p_{i1}^{\text{mis}}$ in the LI scenario so β_{0r} and β_{1r} can be reduced to β_0 and β_1. For MNAR, we assumed the probability of missing compliance data in the treatment arm is related with u_{i1}.

The intercept terms were chosen to control the expected missingness probability at about 0.5 in each scenario. Under the LI assumption, we set $\beta_1 = 2$, referring to the scenario in which the missingness probabilities depend on the probability of being a complier. In a study with a higher proportion of compliers, the noncompliance rates tend to be smaller such that the ITT analysis would perform well. Thus, it may imply a higher probability of not reporting compliance information. Thus the coefficient β_1 was set to be a positive value, matching the above situation.

Under the MNAR assumption, we set $\gamma_{11} = -2$ to produce a scenario in which the missingness in the treatment arm is related to the response rate in the compliers, while the missing probability in the control arm was set to be a fixed value of 0.5. As the true value of θ_i^{CACE} in our setting is negative (a beneficial complier average causal effect if the outcome $o = 1$ is an adverse event), we therefore created a scenario with: 1) a fixed response rate for a control complier (v_{i1}); and 2) a decreasing response rate as u_{i1} increases in the treated complier. Thus, the beneficial CACE tended to be more significant such that it was more likely that the study's investigators would not report compliance information. To do this, we set the coefficient γ_{11} to be negative. The value $\gamma_{11} = -2$ implies a reasonable strength of the association between p_{i1}^{mis} and u_{i1}.

Under each missingness assumption and each true random-effect model, we compare the performance of our proposed method with the naive approach (described in Section 3.2.2) that includes only studies with complete data. Note that to create the missing compliance data, we just added N_{ir11} and N_{ir01} up to create the marginal N_{ir*1}, and added N_{ir10} and N_{ir00} up to give N_{ir*0}. For the analyses using the data from all studies, as no studies were discarded so the true underlying parameters are still in effect, and

the proposed approach can be robust to different missing data generating mechanisms. However, for the naive approach that only includes studies with complete compliance information, patterns of missing mechanism are expected to have an impact on the results.

We used a model selection procedure in the simulation, fitting each dataset with all of the following candidate models: 1) no random effect; 2) random effects only on $(\delta_{in}, \delta_{ia})$; 3) random effects on $(\delta_{in}, \delta_{ia}, \delta_{is})$, and 4) random effects on $(\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu})$, and we counted the frequency of selecting each model using DIC. Note that either π_{ic} or u_{i1} must be generated with a random effect to ensure the missingness probabilities vary across studies. Thus under LI, we generate data with random effects $(\delta_{in}, \delta_{ia})$, $(\delta_{in}, \delta_{ia}, \delta_{is})$, and $(\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu})$, and under MNAR we generate data with $(\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu})$.

3.4.2 Simulation Results

Table 3.4.2 summarizes results from the simulation studies regarding θ^{CACE}, comparing the two approaches in terms of relative bias (ReBias), mean square error (MSE), 95% credible interval coverage probability (CP), 95% credible interval length (CIL), and relative efficiency (RelEff), defined as MSE from the naive analysis including studies with complete data only divided by MSE using the proposed model. Under each missingness mechanism considered, we fit the model including the same random effects as in generating the data. Under the different missingness assumptions, the proposed models were shown to provide nearly unbiased estimates for θ^{CACE} with smaller MSE. Generally, the estimates were slightly biased when the data were generated under LI or MNAR compared to MCAR, or as the number of random effects increased. The coverage probabilities remained close to or above the nominal level 0.95 in all scenarios. The naive method that discards studies with incomplete data also performed reasonably well with no or trivial bias. When data were generated under the MCAR or LI missingness mechanism, though our proposed method was more efficient with consistently smaller MSE and shorter 95% credible interval length, as it gained efficiency by including information from more studies. However, when data were generated with missingness probabilities that were strongly associated with one component of θ_i^{CACE} (the MNAR assumption),

the naive approach using only studies with complete compliance data had substantially large relative bias and MSE. Moreover, the relative efficiency values were greater than two in all scenarios, providing evidence that our proposed model is much more efficient than simply discarding studies without complete compliance data.

Table 3.6: Simulation results: relative bias (ReBias), mean square error (MSE), 95% CI coverage probabilities (CP), and 95% credible interval length (CIL) for θ^{CACE}

Missing	Random	Model inclucing all studies				On studies with complete data				ReEff
Mechanism	Effects	ReBias	MSE	CP	CIL	ReBias	MSE	CP	CIL	
MCAR	None	0.003	0.001	0.951	0.106	0.006	0.003	0.957	0.202	3.599
	δ_{in}, δ_{ia}	0.017	0.001	0.969	0.133	0.018	0.002	0.960	0.198	2.294
	$\delta_{in}, \delta_{ia}, \delta_{is}$	-0.013	0.001	0.953	0.134	-0.001	0.003	0.952	0.200	2.432
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	0.010	0.002	0.988	0.240	-0.087	0.007	0.993	0.455	3.022
LI, $\beta_1 = 2$	δ_{in}, δ_{ia}	0.046	0.001	0.950	0.142	0.008	0.003	0.957	0.219	2.204
	$\delta_{in}, \delta_{ia}, \delta_{is}$	0.030	0.001	0.963	0.142	0.003	0.003	0.951	0.217	2.627
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	0.065	0.004	0.961	0.247	-0.066	0.008	0.989	0.445	2.321
MNAR, $\gamma_{11} = -2$	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	-0.022	0.003	0.978	0.241	-0.319	0.021	0.945	0.507	7.493

ReBias = Bias/True Value

Table 3.7: The estimated probability of selecting a candidate model as the final model using DIC, based on simulation studies with 2000 replicates

Missing Mechanism	True Random Effects Model	Selected Random Effects Model			
		None	δ_{in}, δ_{ia}	$\delta_{in}, \delta_{ia}, \delta_{is}$	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$
MCAR	None	**98.7**	1.2	0.1	0
	δ_{in}, δ_{ia}	0	**94.6**	4.85	0.55
	$\delta_{in}, \delta_{ia}, \delta_{is}$	0	2.8	**95.0**	2.2
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	0	0.05	1.0	**98.95**
LI, $\beta_1 = 2$	δ_{in}, δ_{ia}	0	**95.65**	4.2	0.15
	$\delta_{in}, \delta_{ia}, \delta_{is}$	0	2.9	**94.9**	2.2
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	0	0.4	1.8	**97.8**
MNAR, $\gamma_{11} = -2$	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	0	0.02	1.45	**98.35**

Bolded cell are the probability of identifying the correct model. The probabilities have been multiplied by 100 for presentation. MCAR: missing completely at random, LI: latent ignorable, MNAR: missing not at random.

Under each missing mechanism and true data generating model, we fit four candidate models with different numbers of random effects as described in Section 3.4.1. Table 3.4.2 summarizes the frequency of selecting each candidate model as the "best" model in each set of simulations. For all data-generating scenarios, DIC had a probability around 0.95 of identifying the true random effects model. The results indicate that the Bayesian hierarchical models account for the uncertainty in estimation and produce an appropriate selection of random effects even under the MNAR missing mechanism.

Table 3.8: Performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets

Missing Mechanism	Random Effects Generating Data		Random Effects in Models			
			None	δ_{in}, δ_{ia}	$\delta_{in}, \delta_{ia}, \delta_{is}$	$\delta_{in}, \delta_{ia}, \delta_{ia}, \delta_{iu}$
MCAR	None	ReBias	**0.006**	0.005	-0.021	0.023
		MSE	**0.001**	0.001	0.001	0.001
		CP*	**0.953**	0.972	0.968	0.998
		CIL**	**0.106**	0.129	0.126	0.210
	δ_{in}, δ_{ia}	ReBias	0.012	**0.005**	-0.020	0.025
		MSE	0.002	**0.001**	0.001	0.001
		CP	0.847	**0.961**	0.955	0.997
		CIL	0.110	**0.133**	0.132	0.220
	$\delta_{in}, \delta_{ia}, \delta_{is}$	ReBias	0.010	-0.007	**-0.013**	0.035
		MSE	0.002	0.001	**0.001**	0.002
		CP	0.836	0.922	**0.956**	0.996
		CIL	0.110	0.132	**0.133**	0.221
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	ReBias	0.012	0.086	0.097	**0.011**
		MSE	0.003	0.006	0.006	**0.003**
		CP	0.680	0.637	0.664	**0.978**
		CIL	0.110	0.142	0.148	**0.239**
LI, $\beta_1 = 2$	δ_{in}, δ_{ia}	ReBias	0.067	**0.048**	0.020	0.093
		MSE	0.002	**0.002**	0.001	0.003
		CP	0.795	**0.938**	0.959	0.978
		CIL	0.118	**0.143**	0.141	0.231
	$\delta_{in}, \delta_{ia}, \delta_{is}$	ReBias	0.066	0.040	**0.031**	0.101
		MSE	0.002	0.002	**0.001**	0.003
		CP	0.799	0.920	**0.959**	0.979
		CIL	0.118	0.142	**0.142**	0.231
	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	ReBias	0.075	0.134	0.147	**0.071**
		MSE	0.004	0.007	0.007	**0.004**
		CP	0.669	0.620	0.625	**0.966**
		CIL	0.118	0.152	0.158	**0.248**
MNAR, $\gamma_{11} = -2$	$\delta_{in}, \delta_{ia}, \delta_{is}, \delta_{iu}$	ReBias	0.001	-0.071	-0.029	**-0.022**
		MSE	0.003	0.004	0.004	**0.003**
		CP	0.668	0.671	0.748	**0.983**
		CIL	0.111	0.129	0.139	**0.241**

Bold-face cells are the correct model. **CP: 95% credible interval coverage probability. *CIL: 95% equal-tail credible interval length.

Table 3.4.2 shows the relative bias (ReBias), mean square error (MSE), 95% credible interval coverage probability (CP), and 95% credible interval length (CIL) for θ^{CACE} fitting the four candidate models under each data generating scenario. We present this table to check the performance of our proposed method when fitting the correct model, as well as models that over-fit (including a random effect when it is not present), and under-fit (omitting a random effect when it is present). Regardless of the missing mechanism, when data were fit by a model with the right random effects, the estimated θ^{CACE} generally had a small relative bias, and the coverage probability was around the 95% nominal level. In general, over-fitting produced slightly biased point estimates of θ^{CACE}. Also, 95% credible intervals tended to become too wide as more random effects were included in the model. This is especially true when δ_{iu} was added, which was expected because δ_{iu} directly adds variability to θ_i^{CACE}. Regarding the 95% credible interval coverage probabilities: 1) under the correct model or an over-fitting model, the coverage probabilities were close to or greater than the nominal 0.95; and 2) under an under-fitting model, failure to include the random effect for the study-specific response rate of a never-taker (δ_{is}) did not affect the coverage probability notably, but failure to include the random effect for the probability of the latent compliance groups (δ_{in} or δ_{ia}), or the study-specific response rate of a treated complier (δ_{iu}) reduced coverage for θ^{CACE} substantially.

The results indicated that random effects should be selected carefully to account for potential between-study heterogeneity when estimating CACE in a meta-analysis. For the model performance in terms of the posterior estimates of θ^{CACE} under different missing assumptions, no notable differences were identified, which was expected because of the way we generated missing compliance data. As noted in Section 3.4.1, we did not discard any study under either LI or MNAR, but just combined the generated data from corresponding table cells with affected compliance status to create marginal total counts. This did not affect the point estimates of parameters though it did reduce the efficiency of estimates produced by our proposed models.

3.5 Discussion

We proposed a Bayesian hierarchical model to estimate CACE in meta-analysis of RCTs, accounting for both heterogeneous and incomplete noncompliance among studies, and we applied it to a case study of epidural analgesia trials. We conducted simulation studies to evaluate the performance of our approach under different missingness mechanisms, and the impact of misspecification of random effects. To the best of our knowledge, this is the first meta-analysis of RCTs estimating CACE while adjusting for incomplete noncompliance data.

Using the proposed method, all 27 epidural analgesia trials are included in the CACE meta-analysis. Including information from studies with incomplete data may introduce more heterogeneity into the meta-analysis and may affect the CACE estimate. The simulations indicated that 1) our approach had a good chance of identifying the correct model, and 2) our proposed model had better efficiency for estimating CACE, with smaller MSE and shorter credible intervals, compared to the model only using trials with complete compliance data. Our simulations under the MNAR assumption did not discard any studies when fitting the proposed models, so we expected they would still give unbiased CACE estimates. Nevertheless, the naive approach including only studies with complete compliance data gave biased estimates, which was not surprising because the studies with complete data are no long representative for all studies under the MNAR missing mechanism.

Besides handling the situation in which some studies in a meta-analysis do not report compliance information, the proposed method can be extended to handle missing outcomes. Missing outcome data in RCTs commonly happens when researchers do not collect follow-up outcome for some subjects. For example, consider a trial in which patients randomized to the treatment group were encouraged to receive a flu shot, but the patients themselves decided whether to receive flu shots, and their actual treatment received was recorded. For the outcome of flu-related hospitalization, missing outcomes could occur if some patients had the flu but were treated at hospitals not participating in the study, or if some patients simply had unknown hospitalization status. In this

case, we can extend the likelihood in Equation (3.2) by adding a column "missing" to the right of Table 3.2.1, and the corresponding probabilities would be the sum of the probabilities of all cells in that row.

Recently, extensions of models estimating CACE with missing data in a single study have been developed. Specifically, Chen et al. (2009) discussed the identifiability and estimation of CACE under a nonignorable missing mechanism; Peng et al. (2004) proposed an extended general location model to estimate the CACE with missing data in the outcome and in baseline covariates. Estimating CACE with missing data in longitudinal and survival outcomes has also been discussed (Yau and Little, 2001; Yang et al., 2012). These methods have been proposed only for the single-study setting; potential extensions for estimating CACE in meta-analysis await further development. Furthermore, as network meta-analysis expands the scope of a conventional pairwise meta-analysis to simultaneously compare multiple treatments by synthesizing both direct and indirect information (Lumley, 2002; Zhang et al., 2014), extending the CACE meta-analysis methods to network meta-analysis is also a promising future research topic that awaits further exploration.

Chapter 4

Estimating Causal Effect using the Bayesian Method with the R Package BayesCACE

This chapter introduces the R package **BayesCACE**, which performs CACE analysis for binary outcomes in a single study, and meta-analysis with either complete or incomplete noncompliance information. This package is currently available from GitHub at https://github.com/JinchengZ/BayesCACE. It uses Markov chain Monte Carlo (MCMC) methods on the R platform through **JAGS**. **JAGS** is a program for analyzing Bayesian hierarchical models using MCMC simulation, which is available for diverse computer platforms including Windows and Mac OS X. Convergence of the MCMC routine can be assessed by the function outputs. The package also provides functions to make posterior trace plots, density plots, and auto-correlation plots. For meta-analysis, the package provides a forest plot of study-specific CACE estimates with 95% credible intervals as well as the overall CACE estimate, to visually display the causal treatment effect comparisons.

This chapter is organized as follows. The next section defines CACE in mathematical notation that will be used throughout the paper. We also describe the assumptions needed to make the CACE a valid causal effect estimator. Section 4.2 presents an overview of the Bayesian hierarchical models for CACE implemented in the **BayesCACE** package. Section 4.3 illustrates use of the package with a case study example and discusses the output structures. Finally, Section 4.4 gives a short discussion with potential future improvements.

4.1 CACE, assumptions and definition

The CACE is a measure of the causal effect of a treatment or intervention on patients who received it as intended by the original group allocation. It is an unbiased causal effect estimate based on five standard assumptions commonly used in causal inference research. First, it assumes that potential outcomes for each participant are independent of the potential outcomes for other participants, known as the *Stable Unit Treatment Value Assumption (SUTVA)*. Second, it assumes that assignment to treatment is random, so that the proportion of compliers should be the same in the intervention and control groups, thus allowing us to estimate one of the core unobserved parameters needed to derive a CACE estimate. Third, it assumes that treatment assignment has an effect on the outcome only if it changes the actual treatment taken, an assumption known as the *exclusion restriction*. For never-takers, for instance, it assumes that simply being assigned to treatment does not affect their outcomes, as they do not actually receive the treatment assignd to them. Fourth, it assumes that assigning the study treatment to participants in the intervention condition induces at least some participants to receive the treatment, so the compliance rate is not zero. Finally, it assumes there is a monotonic relationship between treatment assignment and treatment receipt, which implies there are no individuals for whom assignment to treatment actually reduces the likelihood of receiving treatment (i.e., no defiers). This assumption reduces the number of compliance types for whom estimates are derived, permitting a properly identified model.

We follow Zhou et al. (2019) and introduce notation both on the individual level and on the study level. Suppose a meta-analysis reviews I two-armed randomized trials, N_i is the number of subjects in the i-th trial, and $i \in \{1, \ldots, I\}$. If the data include a single study only, then $I = 1$ and we can remove the subscript i from all notation.

On the individual level, define notation as follows for subject j in trial i. (1) Let $R_{ij} = r$ index the randomization assignment with $r = 0$ for those randomized to control and $r = 1$ for those randomized to the intervention. (2) Let $T_{ij}^r = t \in \{0, 1\}$ be the indicator of whether the individual received the intervention. This is a *potential* outcome under the randomization assignment $r \in \{0, 1\}$, *i.e.*, what the value of t would be for individual (i, j) if $r = 0$ or $r = 1$, respectively. (3) Let $Y_{ij}^{r,t} = o \in \{0, 1\}$ be the *potential* binary outcome under randomization assignment r and treatment received t. Note that the *exclusion restriction* assumption allows us to define $Y_{ij}^t \equiv Y_{ij}^{r,t}$. (4) The sets of $\{Y_{ij}^{r,t}\}$ and $\{T_{ij}^r\}$ are the *potential* outcome and treatment-received status respectively under possible r and t, but for each subject in a trial, only one of the possible values of each set can be observed. Therefore, we denote the observed response and received treatment variables as Y_{ij} and T_{ij}. (5) We allow $T_{ij} = *$ if the actual received treatment is not recorded. Then let $M_{ij} = m$ be the missing indicator corresponding to whether subject j has actual treatment received status on record ($m = 0$) or missing ($m = 1$). (6) Using these potential outcomes, we can define the compliers and the CACE. Let C_{ij} be the latent compliance class of individual j in trial i, defined as follows:

$$C_{ij} = \begin{cases} 0, & \text{never-taker, if } \ (T_{ij}^0, T_{ij}^1) = (0, 0) \\ 1, & \text{complier, if } \ (T_{ij}^0, T_{ij}^1) = (0, 1) \\ 2, & \text{always-taker, if } \ (T_{ij}^0, T_{ij}^1) = (1, 1) \\ 3, & \text{defier, if } \ (T_{ij}^0, T_{ij}^1) = (1, 0). \end{cases}$$

A subject's compliance status C_{ij} is not observable because in a two-arm trial, only one of T_{ij}^1 and T_{ij}^0 can be observed. Based on the observed randomization group and actual treatment received, the compliance classes can be only partially identified.

Now, the complier average causal effect of the i^{th} trial is the average difference between potential outcomes for compliers. In this case, the CACE in study i is $\theta_i^{\text{CACE}} =$

$E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1)$, where the patients for whom $C_{ij} = 1$ are the compliers.

On the study level, N_{irto} denotes the observed number of individuals in study i, randomization group r, actual received treatment group t, and outcome o. If the compliance status of individual (i, j) is not on record, $T_{ij} = t = *$ so the corresponding count is N_{ir*o}, which is the sum of the two unobserved counts N_{ir0o} and N_{ir1o}.

4.2 Estimating CACE

In this section, the Bayesian hierarchical models used to estimate CACE are briefly described. These models form the basis of the framework proposed by Zhou et al. (2019) and underlie the **BayesCACE** package. Besides the notation defined in Section 4.1, define the following parameters for study i. (1) Let π_{ia} and π_{in} be the probabilities of being an always-taker and a never-taker, respectively. Because defiers are ruled out by the monotonicity assumption introduced in Section 4.1, each trial has at most only three compliance classes. Thus the probability of being a complier in study i is $\pi_{ic} = 1 - \pi_{ia} - \pi_{in}$. (2) Define these response probabilities: u_{i1} for a complier randomized to the treatment group; v_{i1} for a complier randomized to the control/placebo group; s_{i1} for a never-taker; and b_{i1} for an always-taker. Thus for study i, the parameters included in the model are $\boldsymbol{\beta}_i = (\pi_{ia}, \pi_{in}, u_{i1}, v_{i1}, s_{i1}, b_{i1})$. As the outcome is binary, the expected difference between outcomes from the two treatment groups among compliers is just the risk difference between u_{i1} and v_{i1}. Therefore, the CACE defined in Section 4.1 can be written as $\theta_i^{\text{CACE}} = E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1) = u_{i1} - v_{i1}$.

4.2.1 CACE for a single trial with noncompliance

Consider first a single trial with noncompliance, *i.e.*, $I = 1$, so all notation and parameters defined earlier are reduced to the version without subscript i. According to Zhou et al. (2019), each observed N_{rto} has a corresponding probability that can be written in terms of parameters defined earlier (see Table 4.1), where λ is the proportion of assigning the active treatment ($\text{P}(R_j = 1)$), which is usually known in randomized

trials. Thus the vector $(N_{000}, N_{001}, N_{010}, N_{011}, N_{100}, N_{101}, N_{110}, N_{111})$ follows a multinomial distribution with parameters N and $\mathbf{p}$, where $N = \sum N_{rto}$ and the elements of $\mathbf{p}$ are listed in Table 4.1.

Table 4.1: Observed data and probabilities of a single study

Observed	Probabilities
N_{000}	$(1-\lambda)\{\pi_c(1-v_1)+\pi_n(1-s_1)\}$
N_{001}	$(1-\lambda)(\pi_c v_1+\pi_n s_1)$
N_{010}	$(1-\lambda)\pi_a(1-b_1)$
N_{011}	$(1-\lambda)\pi_a b_1$
N_{100}	$\lambda\pi_n(1-s_1)$
N_{101}	$\lambda\pi_n s_1$
N_{110}	$\lambda\{\pi_c(1-u_1)+\pi_a(1-b_1)\}$
N_{111}	$\lambda(\pi_c u_1+\pi_a b_1)$

Therefore, the log likelihood is

$$logL(\boldsymbol{\beta}) = N_{000}log\{\pi_c(1-v_1)+\pi_n(1-s_1)\} + N_{001}\log(\pi_c v_1+\pi_n s_1)\}+$$

$$N_{010}log\{\pi_a(1-b_1) + N_{011}log\{\pi_a b_1\} + N_{100}log\{\pi_n(1-s_1)\}+$$

$$N_{101}log\{\pi_n s_1\} + N_{110}log\{(\pi_c(1-u_1)+\pi_a(1-b_1)\} + N_{111}\log(\pi_c u_1+\pi_a b_1) \quad (4.1)$$

Assigning a vague prior distribution $f(\boldsymbol{\beta})$ to the parameters $\boldsymbol{\beta} = (\pi_a,\ \pi_n,\ u_1,\ v_1,\ s_1,\ b_1)$, by Bayes' theorem the joint posterior distribution is proportional to $L(\boldsymbol{\beta})f(\boldsymbol{\beta})$. Functionals of the posterior distribution can be estimated by Gibbs and Metropolis-Hastings sampling algorithms using the software **JAGS** *via* the **rjags** package in R. The CACE for a single study is $u_1 - v_1$, so the posterior of θ^{CACE} is the posterior of $u_1 - v_1$.

4.2.2 CACE for a meta-analysis with complete compliance information

This section introduces two methods for doing a meta-analysis of the CACE when noncompliance data are reported in each trial.

The two-step approach

As described in Section 4.2.1, using the observed data N_{irto}, θ_i^{CACE} is identified for study i. Therefore, to estimate the population-average CACE in a meta-analysis, intuitively we can combine the study-specific estimates and standard errors using a standard meta-analysis method such as the fixed-effect (Laird and Mosteller, 1990) or random effects model (Hedges and Vevea, 1998; Hedges and Olkin, 2014). We call this a "two-step" approach. As the CACE measure is a risk difference, a transformation may be necessary to ensure that the normal distribution assumption is approximately true. Building upon the well-developed R package **metafor**, various estimators suggested in the literature can be estimated to account for potential between-study heterogeneity in the CACE, *e.g.*, the Hunter-Schmidt estimator, the Hedges estimator, the DerSimonian-Laird estimator, the maximum-likelihood or restricted maximum-likelihood estimator, or the empirical Bayes estimator (Viechtbauer, 2010).

The Bayesian hierarchical model

In a meta-analysis, the CACE can also be estimate using the joint likelihood from the Bayesian hierarchical model. This method is systematically introduced in Zhou et al. (2019). The log likelihood contribution of trial i is given by Equation 4.1 by adding a subscript i to each parameter. Then the log likelihood for all trials in the meta-analysis is $log\mathcal{L}(\boldsymbol{\beta}) = \sum_i logL_i(\boldsymbol{\beta}_i)$. Because the studies are probably not exactly identical in their eligibility criteria, measurement techniques, study quality, *etc.*, differences in methods and sample characteristics may introduce heterogeneity to the meta-analysis. One way to model the heterogeneity is to use a random effects model.

To guarantee the desired properties of study i's latent compliance classes and to

account for possible between-study heterogeneity in the compliance class and response probabilities, we use these transformations:

(1) $\pi_{in} = \frac{\exp(n_i)}{1+\exp(n_i)+\exp(a_i)}, \pi_{ia} = \frac{\exp(a_i)}{1+\exp(n_i)+\exp(a_i)}$, where $n_i = \alpha_n + \delta_{in}, a_i = \alpha_a + \delta_{ia}$, and $(\delta_{in}, \delta_{ia})^T \sim N(0,\ \mathbf{\Sigma}_{ps})$, $\mathbf{\Sigma}_{ps} = \begin{pmatrix} \sigma_n^2 & \rho\sigma_n\sigma_a \\ \rho\sigma_n\sigma_a & \sigma_a^2 \end{pmatrix}$.

(2) $g(s_{i1}) = \alpha_s + \delta_{is},\ g(b_{i1}) = \alpha_b + \delta_{ib},\ g(u_{i1}) = \alpha_u + \delta_{iu},\ g(v_{i1}) = \alpha_v + \delta_{iv}$, where $g(\cdot)$ is a link function such as the logit or probit, $\delta_{is} \sim N(0, \sigma_s^2)$, $\delta_{ib} \sim N(0, \sigma_b^2)$, $\delta_{iu} \sim N(0, \sigma_u^2)$, $\delta_{iv} \sim N(0, \sigma_v^2)$.

Here we allow correlation between n_i and a_i, and assign random effect variables to all parameters. However, if a parameter does not very between trials, it can be modeled as a fixed effect. Let $f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \mathbf{\Sigma_0})$ be the distributions described above of all parameters $\boldsymbol{\beta_i} = (\pi_{ia},\ \pi_{in},\ s_{i1},\ b_{i1},\ u_{i1},\ v_{i1})$, where $\boldsymbol{\beta_0}$ is the vector of mean hyper-parameters (α_n, α_a, α_s, α_b, α_u, α_v), and $\mathbf{\Sigma_0}$ is the diagonal covariance matrix containing $\mathbf{\Sigma}_{ps}$, σ_s^2, σ_b^2, σ_u^2 and σ_v^2. If we specify $f(\boldsymbol{\beta_0})$ and $f(\mathbf{\Sigma_0})$ as the prior distributions for the hyper-parameters, then the joint posterior distribution is proportional to the likelihood times the priors, *i.e.*, $\prod_i L_i(\boldsymbol{\beta_i}) f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \mathbf{\Sigma_0}) f(\boldsymbol{\beta_0}) f(\mathbf{\Sigma_0})$.

As stated at the beginning of Section 4.2, $\theta_i^{\text{CACE}} = u_{i1} - v_{i1}$ for study i, so for the meta-analysis, the overall CACE is $\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}}) = E(u_{i1}) - E(v_{i1})$. When a random effect δ_{iu} or δ_{iv} is not assigned in the model, $E(u_{i1}) = g^{-1}(\alpha_u)$ and $E(v_{i1}) = g^{-1}(\alpha_v)$. Otherwise, $E(u_{i1})$ and $E(v_{i1})$ can be estimated by integrating out the random effects, *e.g.*, $E(u_{i1}) = \int_{-\infty}^{+\infty} g^{-1}(\alpha_u + t)\sigma_u^{-1}\phi(\frac{t}{\sigma_u})dt$, where $\phi(\cdot)$ is the standard Gaussian density. If the function $g(\cdot)$ is the probit link, this expectation has a closed form: $E(u_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}})$. If the link function $g(\cdot)$ is logit, a well-established approximation $E(u_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_u}{\sqrt{1+C^2\sigma_u^2}})$ can be used, where $C = \frac{16\sqrt{3}}{15\pi}$ (Zeger et al., 1988). The above formulas also apply to $E(v_{i1})$, the expected response rate of a complier in the control group.

The two-step approach, stated by Lin and Zeng (2010), can be viewed as asymptotically equivalent to the model using the joint likelihood. However, as the two-step approach requires the whole set of parameters to be estimated independently for each study, the total number of effective parameters tends to be larger than this method, so estimates using the Bayesian hierarchical model are likely to be more efficient.

4.2.3 CACE for meta-analysis with incomplete compliance information

Another advantage of the Bayesian hierarchical model is that it can include trials with incomplete compliance data. Commonly, some trials do not report noncompliance data because study investigators do not collect actual received treatment status for some subjects or simply do not report compliance. The two-step approach needs counts for all of the groups defined by randomized assignment, treatment received, and outcome in order to estimate the study specific θ_i^{CACE}, so using this method, trials with incomplete compliance data are simply excluded, making estimation less efficient and potentially biased.

In Chapter 3 we proposed a comprehensive framework to incorporate both heterogeneous and incomplete noncompliance data for estimating the CACE in a meta-analysis of RCTs. Here we present the data structure needed for binary outcomes. Table 4.2 shows the probabilities corresponding to the observed counts data. For study i, randomization group $r \in \{0, 1\}$, if the compliance information is reported, then values of N_{ir0o} and N_{ir1o} are reported, where $o \in \{0, 1\}$, so we assign the marginal count $N_{ir*o} = 0$. Otherwise, we do not have data on outcomes for groups defined by actually received treatment, so only the marginal N_{ir*o} is observed, where N_{ir*o} is the number of patients randomized to treatment arm r who had outcome o, for $r, o \in \{0, 1\}$. In this situation, the two unobserved counts N_{ir0o} and N_{ir1o} are assigned as 0.

Table 4.2: Observed data and probabilities in study i of a meta-analysis with incomplete compliance

Observed	Probabilities
N_{i000}	$(1-\lambda_i)\{\pi_{ic}(1-v_{i1})+\pi_{in}(1-s_{i1})\}$
N_{i001}	$(1-\lambda_i)(\pi_{ic}v_{i1}+\pi_{in}s_{i1})$
N_{i010}	$(1-\lambda_i)\pi_{ia}(1-b_{i1})$
N_{i011}	$(1-\lambda_i)\pi_{ia}b_{i1}$
N_{i100}	$\lambda_i\pi_{in}(1-s_{i1})$
N_{i101}	$\lambda_i\pi_{in}s_{i1}$
N_{i110}	$\lambda_i\{(\pi_{ic}(1-u_{i1})+\pi_{ia}(1-b_{i1})\}$
N_{i111}	$\lambda_i(\pi_{ic}u_{i1}+\pi_{ia}b_{i1})$
N_{i0*0}	$(1-\lambda_i)\{\pi_{ic}(1-v_{i1})+\pi_{in}(1-s_{i1})+\pi_{ia}(1-b_{i1})\}$
N_{i0*1}	$(1-\lambda_i)(\pi_{ic}v_{i1}+\pi_{in}s_{i1}+\pi_{ia}b_{i1})$
N_{i1*0}	$\lambda_i\{(\pi_{ic}(1-u_{i1})+\pi_{ia}(1-b_{i1})+\pi_{in}(1-s_{i1})\}$
N_{i1*1}	$\lambda_i(\pi_{ic}u_{i1}+\pi_{ia}b_{i1}+\pi_{in}s_{i1})$

After organizing the observed data as above, Table 4.2 shows the relation between each observed count and the corresponding probability, which is a function of the parameters defined in Section 4.2.2. As before, λ_i is the known allocation ratio for study i, *i.e*, $\lambda_i = \mathrm{P}(R_{ij}=1)$.

The log likelihood contribution for trial i is obtained from the multinomial distribution:

$$logL_i(\boldsymbol{\beta_i}) = N_{i000}log\{\pi_{ic}(1-v_{i1})+\pi_{in}(1-s_{i1})\} + N_{i001}\log(\pi_{ic}v_{i1}+\pi_{in}s_{i1})+$$

$$N_{i010}log\{\pi_{iia}(1-b_{i1})\} + N_{i011}log\{\pi_{ia}b_{i1}\} + N_{i100}log\{\pi_{in}(1-s_{i1})\}+$$

$$N_{i101}log\{\pi_{in}s_{i1}\} + N_{i110}log\{(\pi_{ic}(1-u_{i1})+\pi_{ia}(1-b_{i1})\} + N_{i111}\log(\pi_{ic}u_{i1}+\pi_{ia}b_{i1})+$$

$$N_{i0*0}log\{\pi_{ic}(1-v_{i1})+\pi_{in}(1-s_{i1})+\pi_{ia}(1-b_{i1})\} + N_{i0*1}\log(\pi_{ic}v_{i1}+\pi_{in}s_{i1}+\pi_{ia}b_{i1})+$$

$$N_{i1*0}log\{(\pi_{ic}(1-u_{i1})+\pi_{ia}(1-b_{i1})+\pi_{in}(1-s_{i1})\}+N_{i1*1}\log(\pi_{ic}u_{i1}+\pi_{ia}b_{i1}+\pi_{in}s_{i1}) \tag{4.2}$$

Because the parameters $\boldsymbol{\beta_i} = (\pi_{ia},\ \pi_{in},\ s_{i1},\ b_{i1},\ u_{i1},\ v_{i1})$ are the same as in Section 4.2.2, the estimation process is also the same: assign distributions $f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \boldsymbol{\Sigma_0})$, where $\boldsymbol{\beta_0}$ is the vector of mean hyper-parameters, and $\boldsymbol{\Sigma_0}$ is the covariance matrix; then specify prior distributions for $f(\boldsymbol{\beta_0})$ and $f(\boldsymbol{\Sigma_0})$, so the joint posterior is proportional to $\prod_i L_i(\boldsymbol{\beta_i}) f(\boldsymbol{\beta_i}|\boldsymbol{\beta_0}, \boldsymbol{\Sigma_0}) f(\boldsymbol{\beta_0}) f(\boldsymbol{\Sigma_0})$. Similarly, the CACE for this meta-analysis incorporating incomplete compliance data is $\theta^{\text{CACE}} = E(\theta_i^{\text{CACE}}) = E(u_{i1}) - E(v_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}}) - \Phi(\frac{\alpha_v}{\sqrt{1+\sigma_v^2}})$ if the probit link function is used for u_{i1} and v_{i1}.

4.3 Using the R package BayesCACE

The primary objective of the **BayesCACE** package is to provide a user-friendly implementation of the Bayesian method for estimating the CACE, described in Section 4.2. The package is now available to download and install from GitHub at `https://github.com/JinchengZ/BayesCACE`. It will soon be available via the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=BayesCACE`. The **BayesCACE** package depends on the R packages **rjags** (Plummer, 2018), **coda** (Plummer et al., 2006), and **forestplot** (Gordon and Lumley, 2017). Users need to install **JAGS** separately from its homepage `http://mcmc-jags.sourceforge.net` as the **BayesCACE** package does not include a copy of the **JAGS** library. The current version of **JAGS** is 4.3.0, which is the version of the package **BayesCACE** requires; earlier versions of **JAGS** may not guarantee exactly reproducible results. Once the package has been correctly installed, it should be possible to replicate the analyses described in this section to within MCMC error.

4.3.1 Data structure for estimating the CACE

We introduce the data structures through the illustration example included in the package **BayesCACE**: `epidural_c` and `epidural_ic`. These two data sets were obtained from Bannister-Tyrrell et al. (2015), who conducted an exploratory meta-analysis of

the association between using epidural analgesia in labor and the risk of cesarean section. Data `epidural_c` contains 10 trials with full compliance information; each trial has 8 observed counts, denoted by N_{irto} and presented in columns `Nirto`, $i = 1, \ldots, 10$, $r, t, o \in \{0, 1\}$. These data were re-analyzed by Zhou et al. (2019) in a meta-analysis using their proposed Bayesian hierarchical model to estimate the CACE. The function `cace.meta.c()` introduced in Section 4.3.3 performs this analysis. The column `study.id` contains IDs for the 10 studies, and `study.name` labels each study by its first author's surname and its publication year.

The data can be loaded and printed using these commands:

```
R> library("BayesCACE")
R> data("epidural_c", package = "BayesCACE")
R> epidural_c
```

	study.id	study.name	N000	N001	N010	N011	N100	N101	N110	N111
1	1	Bofill, 1997	37	2	11	1	2	0	42	5
2	2	Clark, 1998	72	6	68	16	7	2	134	13
3	3	Halpern, 2004	62	5	44	7	0	0	112	12
4	4	Head, 2002	51	7	2	0	3	0	43	10
5	5	Jain, 2003	72	11	0	0	0	2	36	7
6	6	Nafisi, 2006	179	19	0	0	0	0	173	24
7	7	Nikkola, 1997	6	0	4	0	0	0	10	0
8	8	Ramin, 1995	546	17	95	8	230	2	393	39
9	9	Sharma, 1997	336	16	5	0	114	1	231	12
10	10	Volmanen, 2008	23	1	3	0	1	0	23	1

The other dataset `epidural_ic` represents the situation in which not all trials report complete compliance data. It contains 27 studies, only 10 out of which have full compliance information and were included in `epidural_c`. This dataset is also drawn from Bannister-Tyrrell et al. (2015) but only the method introduced in Section 4.2.3 can include the studies with incomplete compliance information when estimating the CACE. The function `cace.meta.ic()` (see Section 4.3.3 for details) performs this analysis.

Each study is represented by one row in the dataset; the columns `study.id` and `study.name` have the same meanings as in the dataset `epidural_c`. Each study's data is summarized in 12 numbers (columns) denoted by N_{irto} and N_{ir*o} as described in Section 4.2.3. For a particular randomization group $r \in \{0, 1\}$, observed counts are presented either as N_{irto} or N_{ir*o} depending on whether the compliance information is available; values for other columns are denoted by 0. The corresponding column names in the dataset are `Nirto` and `Nirso`, respectively.

The first 6 rows of the dataset `epidural_ic` are printed below.

```
R> data("epidural_ic", package = "BayesCACE")
R> head(epidural_ic)
```

	study.id	study.name	N000	N001	N010	N011	N0s0	N0s1	N100	N101	N110	N111	N1s0	N1s1
1	1	Bofill, 1997	37	2	11	1	0	0	2	0	42	5	0	0
2	2	Clark, 1998	72	6	68	16	0	0	7	2	134	13	0	0
3	3	Dickinson, 2002	0	0	0	0	428	71	0	0	0	0	408	85
4	4	Evron, 2008	40	4	0	0	0	0	0	0	0	0	129	19
5	5	El Kerdawy, 2010	0	0	0	0	12	3	0	0	0	0	11	4
6	6	Gambling, 1998	0	0	0	0	573	34	206	10	371	29	0	0

Note that `NA` is not allowed in a dataset for the package **BayesCACE**, but some trials may have 0 events or 0 noncompliance rates.

4.3.2 Plotting noncompliance rates

Before doing the CACE analysis, one might want a visual overview of the study-specific noncompliance rates in both randomization arms. The function `plot.noncomp` provides a forest plot of noncompliance rates in an R plot window. The function can be simply called as

```
plot.noncomp(data, overall=TRUE),
```

where `data` is a dataset with structure like `epidural_c` or `epidural_ic`. Only studies with full compliance information are included in this plot because noncompliance

rates cannot be calculated without compliance data. Figure 4.1 shows the resulting plot, where the red dot with its horizontal line shows the study-specific noncompliance rate with its 95% exact confidence interval for the patients randomized to the treatment arm, and the blue square with its horizontal line represents that rate and interval for those in the control arm. The confidence intervals are calculated by the Clopper-Pearson exact method (Clopper and Pearson, 1934), which is based on the cumulative distribution function of the binomial distribution. By default `overall=TRUE` so that the figure also gives a summary estimate of the compliance rates per randomization group. This overall rate is estimated using a logit generalized linear mixed model. Otherwise if the argument `overall` is `FALSE`, the plot shows only study-specific noncompliance rates.

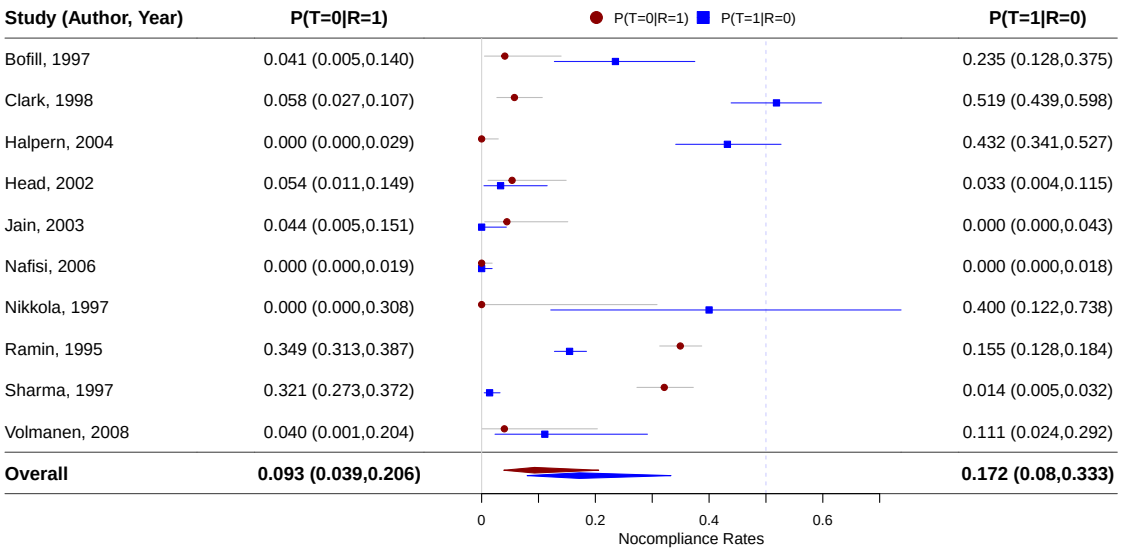

Figure 4.1: Noncompliance rates plot generated by the function `plot.noncomp()`.

4.3.3 CACE analysis for a single study or in a meta-analysis

The major functions in the **BayesCACE** package are `cace.study()`, `cace.meta.c()`, and `cace.meta.ic()`, which implement the models introduced in Section 4.2 to perform Bayesian CACE analysis for different data structures. In particular, `cace.study()` does

CACE analysis for a single study using the likelihood and model specified in Section 4.2.1. The function `cace.meta.c()` does CACE analysis for a meta-analysis when each trial reports noncompliance information. Users can choose to do the analysis either by the two-step approach or using the Bayesian hierarchical model, as introduced in Section 4.2.2. When some trials do not report noncompliance data, the function `cace.meta.ic()` can be applied to perform a CACE meta-analysis using the likelihood in Equation 4.2. The commands in each function may take 1-15 minutes to run. Generally the two-step approach using the function `cace.meta.c()` takes longer because MCMC chains are run on the studies one by one. The actual run time depends on the amount of data and the user's processor.

Function `cace.study()` for a study-specific analysis or a two-step meta-analysis

For the default interface, the arguments of the function `cace.study()` are

```
cace.study(data, param = c("CACE", "u1", "v1", "s1", "b1", "pi.c", "pi.n",
    "pi.a"), prior.type = "default", digits = 3, n.adapt = 1000, n.iter =
    100000, n.burnin = floor(n.iter/2), n.chains = 3, n.thin = max(1, floor
    ((n.iter - n.burnin)/1e+05)), conv.diag = FALSE, mcmc.samples = FALSE,
    two.step = FALSE, method = "REML")
```

where users need to input `data` with the same structure as `epidural_c`, containing either one row of observations for a single study, or multiple rows referring to multiple studies in a meta-analysis. This function fits a model for a single study as described in Section 4.2.1. If the data includes more than one study, the study-specific CACEs will be estimated by retrieving data row by row.

The argument `param` is a character string vector indicating the parameters to be tracked and estimated. By default all parameters shown in Section 4.2.1 are included: θ^{CACE} (`CACE`), u_1 (`u1`), v_1 (`v1`), s_1 (`s1`), b_1 (`b1`), π_a (`pi.a`), π_n (`pi.n`), and $\pi_c = 1-\pi_a-\pi_n$ (`pi.c`). Users can modify the string vector to only include parameters of interest besides θ^{CACE}. If users do not specify their own prior distributions, the default priors are used (`prior.type = "default"`). They are assigned to the transformed scale

of the following parameters: $\pi_n = \frac{\exp(n)}{1+\exp(n)+\exp(a)}, \pi_a = \frac{\exp(a)}{1+\exp(n)+\exp(a)}$, $\text{logit}(s_1) = \alpha_s$, $\text{logit}(b_1) = \alpha_b$, $\text{probit}(u_1) = \alpha_u$, $\text{probit}(v_1) = \alpha_v$, where $n, a \sim N(0, 2.5^2)$ and $\alpha_s, \alpha_b, \alpha_u, \alpha_v \sim N(0, 2^2)$. With these settings, a 95% prior probability interval for any of the probabilities π_{in}, π_{ia}, and π_{ic} ranges from about 0.001 to 0.91, and a 95% prior interval for the probabilities s_1, b_1, u_1, and v_1 ranges from about 0.01 to 0.98. Alternatively, users can specify their own prior distributions for all parameters, and save them as a file `prior.study.R` under the same directory with the model function. By assigning `prior.type = "custom"`, the function calls the user-defined text string as the priors. Note that if users choose the customized priors, the pre-defined `prior.study.R` must include distributions for all parameters. The function cannot combine the default priors with partial user-defined prior distributions. Here we give an example of `prior.study.R` when assigning $N(0, 10^2)$ to every parameter:

```
prior.study <- function(prior.type="custom"){
  string2 <-
  "# priors
  n ~ dnorm(0, 0.01)
  a ~ dnorm(0, 0.01)
  alpha.s ~ dnorm(0, 0.01)
  alpha.b ~ dnorm(0, 0.01)
  alpha.u ~ dnorm(0, 0.01)
  alpha.v ~ dnorm(0, 0.01)
  }"
return(string2)
}
```

The arguments `n.adapt`, `n.iter`, `n.burnin`, `n.chains`, and `n.thin` control the MCMC algorithm run by the R package **rjags** (Plummer, 2018). The argument `n.adapt` is the number of iterations for adaptation; it is used to maximize the sampling efficiency, and the default is set as 1,000. The argument `n.chains` determines the number of MCMC chains (the default is 3); `n.iter` is the number of iterations of each MCMC

chain; `n.burnin` is the number of burn-in iterations at the beginning of each chain to be discarded; `n.thin` is the thinning rate for MCMC chains, which is used to avoid potential high auto-correlation and to save computer memory when `n.iter` is large. The default of `n.thin` is set as `1` or the largest integer not greater than `((n.iter-n.burnin)/1e+05))`, whichever is larger. The argument `conv.diag` specifies whether to compute the Gelman and Rubin convergence statistic ($\hat{R}$) of each parameter as a convergence diagnostic (Brooks and Gelman, 1998; Gelman and Rubin, 1992). It is considered the chains are well mixed and have converged to the target distribution if $\hat{R} \leq 1.1$. If the argument `mcmc.samples = TRUE`, the function saves each chain's MCMC samples for all parameters, which can be used to produce trace, posterior density, and auto-correlation plots by calling the function `plot.cacebayes`.

By default, the function `cace.study()` returns a list including posterior estimates (posterior mean, standard deviation, median, and a 95% credible interval (CI) with 2.5% and 97.5% quantiles as the lower and upper bounds), and the deviance information criterion (DIC) statistic (Spiegelhalter et al., 2002) for each study. The argument `two.step` is a logical value indicating whether to conduct a two-step meta-analysis. If `two.step = TRUE`, the posterior mean and standard deviation of study-specific θ_i^{CACE} are used to do a standard meta-analysis, using the R package **metafor**. The default estimation method is the REML (restricted maximum-likelihood estimator) method for the random-effects model (Harville, 1977). Users can change the argument `method` to obtain different meta-analysis estimators from either a random-effects model or a fixed-effect model, e.g., `method = "DL"` refers to the DerSimonian-Laird estimator, `method = "HE"` returns the Hedges estimator, and `method = "HS"` gives the Hunter-Schmidt estimator. More details are available from the documentation of the function `metafor::rma` (Viechtbauer, 2010). If the input data include only one study, the meta-analysis result is just the same as the result from the single study.

Here is an example to demonstrate the function's usage. We call the function `cace.study()` on the dataset `epidural_c` as follows:

```
R> data("epidural_c", package = "BayesCACE")
R> set.seed(123)
```

```
R> out.study <- cace.study(data = epidural_c, conv.diag = TRUE, mcmc.samples
+   = TRUE, two.step = TRUE)
```

The following messages are output as the code runs:

```
NA is not allowed in the input data set;
the rows containing NA are removed.
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 2
   Unobserved stochastic nodes: 6
   Total graph size: 44

Initializing model

  |++++++++++++++++++++++++++++++++++++++++++++++++++| 100%
  |**************************************************| 100%
  |**************************************************| 100%
MCMC convergence diagnostic statistics are calculated and saved in conv.out
```

If the dataset contains more than one study, e.g., the `epidural_c` dataset has 10 trials, then once the JAGS model compiles for the first study, it automatically continues to run on the next study's data. The results are saved in the object `out.study`, a list containing the model name, posterior information for each monitored parameter, and DIC of each study. We can use parameter names to display the corresponding estimates. The argument `digits` in the function `cace.study()` can be used to change the number of significant digits to the right of the decimal point. Here, we used the default setting `digits = 3`. For example, the estimates of θ^{CACE} for each single study (posterior mean and standard deviation, posterior median, 95% credible interval, and time-series standard error) can be displayed as

```
R> out.study$CACE
```

```
          Mean     SD    2.5%       50%  97.5% Time-series SE
 [1,]  0.04960 0.0796 -0.0944  4.41e-02 0.2180 2.52e-04
 [2,] -0.02460 0.0488 -0.1220 -2.19e-02 0.0789 1.48e-04
 [3,] -0.02180 0.0609 -0.1270 -2.88e-02 0.1130 1.93e-04
 [4,]  0.07180 0.0762 -0.0769  7.12e-02 0.2240 2.05e-04
 [5,]  0.08260 0.0765 -0.0620  8.13e-02 0.2370 2.52e-04
 [6,]  0.02600 0.0318 -0.0362  2.58e-02 0.0887 7.42e-05
 [7,]  0.01420 0.1560 -0.2770  2.11e-04 0.4000 4.07e-04
 [8,]  0.05020 0.0247  0.0024  5.00e-02 0.0992 7.26e-05
 [9,] -0.01090 0.0234 -0.0571 -1.08e-02 0.0349 6.29e-05
[10,]  0.00127 0.0649 -0.1340 -3.87e-06 0.1430 1.53e-04
```

If the argument `conv.diag` is specified as `TRUE`, the output list contains a sub-list `conv.out`, which outputs the point estimates of the 'potential scale reduction factor' (the Gelman and Rubin convergence statistic, labelled `Point est.`) calculated for each parameter from each single study, and their upper confidence limits (labelled `Upper C.I.`). Approximate convergence is diagnosed when the upper limit is close to 1 (Brooks and Gelman, 1998; Gelman and Rubin, 1992). For example, the first sub-list from `conv.out` is

```
R> out.study$conv.out[[1]]
```

```
     Point est. Upper C.I.
CACE  1.0000007   1.000003
b1    1.0000224   1.000060
pi.a  1.0000338   1.000127
pi.c  1.0000380   1.000135
pi.n  1.0000148   1.000063
s1    1.0000135   1.000042
u1    1.0000121   1.000028
v1    0.9999995   1.000012
```

Also, in this example, we included `mcmc.samples = TRUE` in the argument, so the output object list `out.study` includes each chain's MCMC samples for all parameters. They can be used in the function `plot.cacebayes` to generate the trace, posterior density, and auto-correlation plots for further model diagnostics.

If the dataset used by the function `cace.study()` has more than one study, specifying the argument `two.step = TRUE` causes the two-step meta-analysis for θ^{CACE} to be done. The outcomes are saved as a sub-list object `meta`. Note that users can obtain different meta-analysis estimators by changing the `method` argument as described earlier.

```
R> out.study$meta

Random-Effects Model (k = 10; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.0002 (SE = 0.0008)
tau (square root of estimated tau^2 value):      0.0129
I^2 (total heterogeneity / total variability):   8.00%
H^2 (total variability / sampling variability):  1.09

Test for Heterogeneity:
Q(df = 9) = 5.9134, p-val = 0.7486

Model Results:

estimate      se    zval    pval    ci.lb   ci.ub
  0.0183  0.0142  1.2854  0.1986  -0.0096  0.0462

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Function `cace.meta.c()` for meta-analysis with complete compliance data

The function `cace.meta.c()` performs the Bayesian hierarchical model method for meta-analysis when the dataset has complete compliance information for all studies, as described in Section 4.2.2. The function's default arguments are given by

```
cace.meta.c(data, param = c("CACE", "u1out", "v1out", "s1out", "b1out",
    "pic", "pin", "pia"), prior.type = "default", delta.n = TRUE, delta.a
    = TRUE, delta.u = TRUE, delta.v = TRUE, delta.s = TRUE, delta.b = TRUE,
    cor = TRUE, digits = 3, n.adapt = 1000, n.iter = 100000, n.burnin =
    floor(n.iter/2), n.chains = 3, n.thin = max(1, floor((n.iter - n.burnin)
    /100000)), conv.diag = FALSE, mcmc.samples = FALSE,
    study.specific = FALSE)
```

The arguments controlling the MCMC algorithm are mostly similar to those of `cace.study()`. One major difference is that users need to specify parameters that are modeled as random effects. In Section 4.2.2, we showed how to specify random effects for each parameter on the transformed scales, namely δ_{in}, δ_{ia}, δ_{iu}, δ_{iv}, δ_{is}, and δ_{ib}, and allowed a non-zero correlation ρ between δ_{in} and δ_{ia}. The model with all of these random effects as well as the correlation ρ is considered the full model. However, this function is flexible, allowing users to choose which random effects to include by specifying the logical-valued arguments `delta.n`, `delta.a`, `delta.u`, `delta.v`, `delta.s`, `delta.b`, and `cor`, respectively. The default model sets all of these arguments to `TRUE`. Note that ρ (`cor`) can only be included when both δ_{in} (`delta.n`) and δ_{ia} (`delta.a`) are set to `TRUE`. Otherwise, a warning occurs and the model continues running by forcing `delta.n = TRUE` and `delta.a = TRUE`. The default parameters to be monitored depend on which parameters are modeled as random effects. For example, `u1out` refers to $E(u_{i1})$ as described in Section 4.2.2, where for the probit link, $E(u_{i1}) = \Phi(\alpha_u)$ if δ_u is not specified in the model, and $E(u_{i1}) = \Phi(\frac{\alpha_u}{\sqrt{1+\sigma_u^2}})$ when the random effect δ_u is included.

The argument `prior.type = "default"` is used if users do not specify prior distributions. Like the function `cace.study`, weakly informative priors $\alpha_n, \alpha_a \sim N(0, 2.5^2)$

and $\alpha_s, \alpha_b, \alpha_u, \alpha_v \sim N(0, 2^2)$ are assigned to the means of these transformed parameters: $\pi_{in} = \frac{\exp(n_i)}{1+\exp(n_i)+\exp(a_i)}$, $\pi_{ia} = \frac{\exp(a_i)}{1+\exp(n_i)+\exp(a_i)}$, where $n_i = \alpha_n + \delta_{in}, a_i = \alpha_a + \delta_{ia}$; $\text{logit}(s_{i1}) = \alpha_s + \delta_{is}$, $\text{logit}(b_{i1}) = \alpha_b + \delta_{ib}$, $\text{probit}(u_{i1}) = \alpha_u + \delta_{iu}$, and $\text{probit}(v_{i1}) = \alpha_v + \delta_{iv}$. For the random effects, we have $\delta_{is} \sim N(0, \sigma_s^2)$, $\delta_{ib} \sim N(0, \sigma_b^2)$, $\delta_{iu} \sim N(0, \sigma_u^2)$, and $\delta_{iv} \sim N(0, \sigma_v^2)$, as response rates are assumed to be independent between latent classes. A $Gamma(2, 2)$ hyper-prior distribution is assigned to the precision parameters σ_s^{-2}, σ_b^{-2}, σ_u^{-2} and σ_v^{-2}, which corresponds to a 95% interval of $(0.6, 2.9)$ for the corresponding standard deviations, allowing moderate heterogeneity in the response rates. In a reduced model with one of δ_{in} or δ_{ia} set to 0, the prior of the other precision parameter is also assumed to be $Gamma(2, 2)$, which gives moderate heterogeneity for latent compliance classes probabilities, whereas for the full model, $(\delta_{in}, \delta_{ia})^T \sim N(0, \ \mathbf{\Sigma}_{ps})$, the prior for the variance-covariance matrix $\mathbf{\Sigma}_{ps}$ is $InvWishart(I, 3)$, where I is the identity matrix.

Alternatively, this function allows users to specify their own prior distributions by saving a separate R file `prior.meta.R` under the save directory with the model file, and assigning the argument `prior.type = "custom"`. Because the function `cace.meta.c()` is more complicated depending on the choice of random effects, as an illustration we show an example of the customized prior distributions file when assigning `delta.n = TRUE`, `delta.a = FALSE`, `delta.u = TRUE`, `delta.v = FALSE`, `delta.s = TRUE`, `cor = FALSE` to function `cace.meta.c()`.

```
prior.meta <- function(prior.type="custom"){

string2 <-
" delta.n[i] ~ dnorm(0, tau.n)
  delta.u[i] ~ dnorm(0, tau.u)
  delta.s[i] ~ dnorm(0, tau.s)
  }

# priors
```

```
  alpha.n ~  dnorm(0, 0.16)
  alpha.a ~ dnorm(0, 0.16)
  alpha.s ~  dnorm(0, 0.25)
  alpha.b ~  dnorm(0, 0.25)
  alpha.u ~  dnorm(0, 0.25)
  alpha.v ~  dnorm(0, 0.25)

  tau.n ~ dgamma(2, 2)
  sigma.n <- 1/sqrt(tau.n)
  tau.u ~ dgamma(2, 2)
  sigma.u <- 1/sqrt(tau.u)
  u1out <- phi(alpha.u/sqrt(1+sigma.u^2))
  v1out <- phi(alpha.v)
  CACE <- u1out-v1out
  s1out <- ilogit(alpha.s/sqrt(1 + (16^2*3/(15^2*pi^2))*sigma.s^2))
  tau.s ~ dgamma(2, 2)
  sigma.s <- 1/sqrt(tau.s)
  b1out <- ilogit(alpha.b)
  }"

return(string2)
}
```

Users can modify the above customized file `prior.meta.R` to assign their preferred prior distributions. Note that same as the function `cace.study()`, the function cannot combine the default priors with partial user-defined prior distributions. Thus users need to be careful when choosing the customized priors: the pre-defined R file `prior.meta.R` must include distributions for all hyper-parameters.

The `epidural_c` dataset is used as a real-study example:

```
R> data("epidural_c", package = "BayesCACE")
R> set.seed(123)
R> out.meta.c <- cace.meta.c(data = epidural_c, conv.diag = TRUE,
+ mcmc.samples = TRUE, study.specific = TRUE)
```

The usage of arguments `conv.diag` and `mcmc.samples` are the same as for the function `cace.study`. When the argument `study.specific` is specified as `TRUE`, the model will first check the logical status of arguments `delta.u` and `delta.v`. If both are `FALSE`, meaning that neither response rate u_{i1} or v_{i1} is modeled with a random effect, then the study-specific θ_i^{CACE} is the same across studies. The function gives a warning and continues by making `study.specific = FALSE`. Otherwise, the study-specific θ_i^{CACE} are estimated and saved as the parameter `cacei`.

In this example, by calling the object `smry` from the output list `out.meta.c`, posterior estimates (posterior estimates (posterior mean, standard deviation, posterior median, 95% credible interval, and time-series standard error) are displayed.

```
R> out.meta.c$smry
```

	Mean	SD	2.5%	50%	97.5%	Time-series SE
CACE	0.020900	0.0632	-0.10200	1.94e-02	0.1510	7.72e-04
b1out	0.127000	0.0451	0.05930	1.20e-01	0.2340	3.91e-04
cacei[1]	0.044000	0.0678	-0.08140	4.08e-02	0.1870	2.32e-04
cacei[2]	-0.023100	0.0491	-0.11500	-2.50e-02	0.0820	1.84e-04
cacei[3]	-0.007330	0.0566	-0.10900	-1.14e-02	0.1130	2.13e-04
cacei[4]	0.065400	0.0680	-0.06650	6.46e-02	0.2020	1.66e-04
cacei[5]	0.053800	0.0685	-0.07310	5.11e-02	0.1950	2.42e-04
cacei[6]	0.026300	0.0308	-0.03390	2.61e-02	0.0872	6.78e-05
cacei[7]	0.003040	0.0933	-0.18900	6.39e-05	0.2100	3.56e-04
cacei[8]	0.048400	0.0237	0.00215	4.83e-02	0.0953	6.25e-05
cacei[9]	-0.010700	0.0224	-0.05530	-1.06e-02	0.0331	5.58e-05
cacei[10]	0.000278	0.0604	-0.12100	-1.29e-03	0.1290	2.08e-04
pia	0.121000	0.0804	0.02550	1.02e-01	0.3450	4.69e-03

```
pic        0.815000 0.0948  0.55900  8.34e-01 0.9330 5.84e-03
pin        0.064500 0.0401  0.01540  5.59e-02 0.1590 2.32e-03
s1out      0.183000 0.1040  0.04540  1.60e-01 0.4400 8.93e-04
u1out      0.128000 0.0480  0.05540  1.20e-01 0.2430 6.14e-04
v1out      0.107000 0.0406  0.04740  1.00e-01 0.2040 4.61e-04
```

The posterior estimates of θ_i^{CACE} can be used to make a forest plot by calling the function `plot.forest`, which will be introduced in Section 4.3.5.

Users can manually do model selection procedures by including different random effects and comparing DIC from the outputs. DIC and its two components are saved as an object `DIC` in the output list.

```
R> out.meta.c$DIC
```

```
D.bar 204.34102
pD     44.74046
DIC   249.08148
```

`DIC` is the penalized deviance, calculated as the sum of `D.bar` and `pD`, where `D.bar` is the posterior expectation of the deviance, reflecting the model fit, and `pD` reflects the effective number of parameters in the model. `D.bar` is usually lower when more parameters are included in the model, but complex models may lead to overfitting. Thus `DIC` balances the model's fit against the effective number of parameters. Generally a model with smaller DIC is preferred. However, it is difficult to say what constitutes an important improvement in DIC. Following Lunn et al. (2012), we suggest that a reduction of less than 5 is not a substantial improvement. When fitting models to a particular dataset, it is usually uncertain which random effect variables should be included in the model. The function `cace.meta.c()` allows users to specify candidate models with different random effects, and thus to conduct a forward/backward/stepwise model selection procedure to choose the best fitting model.

Function `cace.meta.ic()` for meta-analysis with incomplete compliance information

Another major function in the package **BayesCACE** is `cace.meta.ic()`. It also estimates θ^{CACE} using the Bayesian hierarchcal model but can accommodate studies with incomplete compliance data. The necessary data structure and the likelihood function are presented in Section 4.2.3. The arguments of this function are

```
cace.meta.ic(data, param = c("CACE", "u1out", "v1out", "s1out", "b1out",
    "pic", "pin", "pia"), prior.type = "default", delta.n = TRUE, delta.a
    = TRUE, delta.u = TRUE, delta.v = TRUE, delta.s  = TRUE, delta.b =
    TRUE, cor = TRUE, n.burnin = floor(n.iter/2), digits = 3, n.adapt =
    1000, n.iter = 100000, n.chains = 3, n.thin = max(1, floor((n.iter -
    n.burnin)/100000)), conv.diag = FALSE, mcmc.samples = FALSE,
    study.specific = FALSE)
```

The arguments of `cace.meta.ic()` are mostly similar to those of `cace.meta.c()`, though `cace.meta.ic()` calls a different built-in model file from the package **BayesCACE**. The major difference in using this function is that users need to create a dataset with the same structure as `epidural_ic`. Please check Section 4.3.1 for data preparation details. As for `cace.meta.c()`, users can set their customized prior distributions by saving them as a separate R file `prior.meta.R`. Here we use the `epidural_ic` dataset as an example:

```
R> data("epidural_ic", package = "BayesCACE")
R> set.seed(123)
R> out.meta.ic <- cace.meta.ic(data = epidural_ic, conv.diag = TRUE,
+   mcmc.samples = TRUE, study.specific = TRUE)
```

The results are saved in the object `out.meta.ic`, a list containing posterior estimates for monitored parameters, DIC, convergence diagnostic statistics, and MCMC samples. In this example, the argument `study.specific` is `TRUE`, so the summary for each study-specific θ_i^{CACE} is displayed in the object `out.meta.ic$smry` together with other parameters.

Note that when compiling the JAGS model, the warning "adaptation incomplete" may occasionally occur, indicating that the number of iterations for the adaptation process is not sufficient. The default value of `n.adapt` (the number of iterations for adaptation) is 1,000. This is an initial sampling phase during which the samplers adapt their behavior to maximize their efficiency (e.g., a Metropolis-Hastings random walk algorithm may change its step size) (Plummer, 2018). The "adaptation incomplete" warning indicates the MCMC algorithm may not achieve maximum efficiency, but it generally has little impact on the posterior estimates of the treatment effects. To avoid this warning, users may increase `n.adapt`.

4.3.4 Plotting the trace plot, posterior density, and auto-correlation

When compiling the JAGS models, it is helpful to assess the performance of the MCMC algorithm. The function `plot.cacebaes()` provides diagnostic plots for the MCMC, namely trace plots, auto-correlation plots and kernel density estimation plots. Both trace plots and the auto-correlation plots can be used to examine whether the MCMC chains appear to be drawn from stationary distributions. A posterior density plot for a parameter visually shows the posterior distribution. Users can simply call this function on objects produced by `cace.study()`, `cace.meta.c()`, or `cace.meta.ic()`.

The arguments of this plot function are

```
plot.cacebayes(obj, which = c("trace", "density", "autocorr"), param =
    c("CACE"), trialnumber = 1, ...)
```

Here, we use the objects list obtained from fitting the Bayesian hierarchical model `cace.meta.ic()` in Section 4.3.3 as examples to generate the three plots. To avoid lengthy output we just illustrate how these plots are produced for θ^{CACE}. The relevant code is:

```
R> plot.cacebayes(obj = out.meta.ic)
```

The resulting plots are shown in Figure 4.2, 4.3, and 4.4.

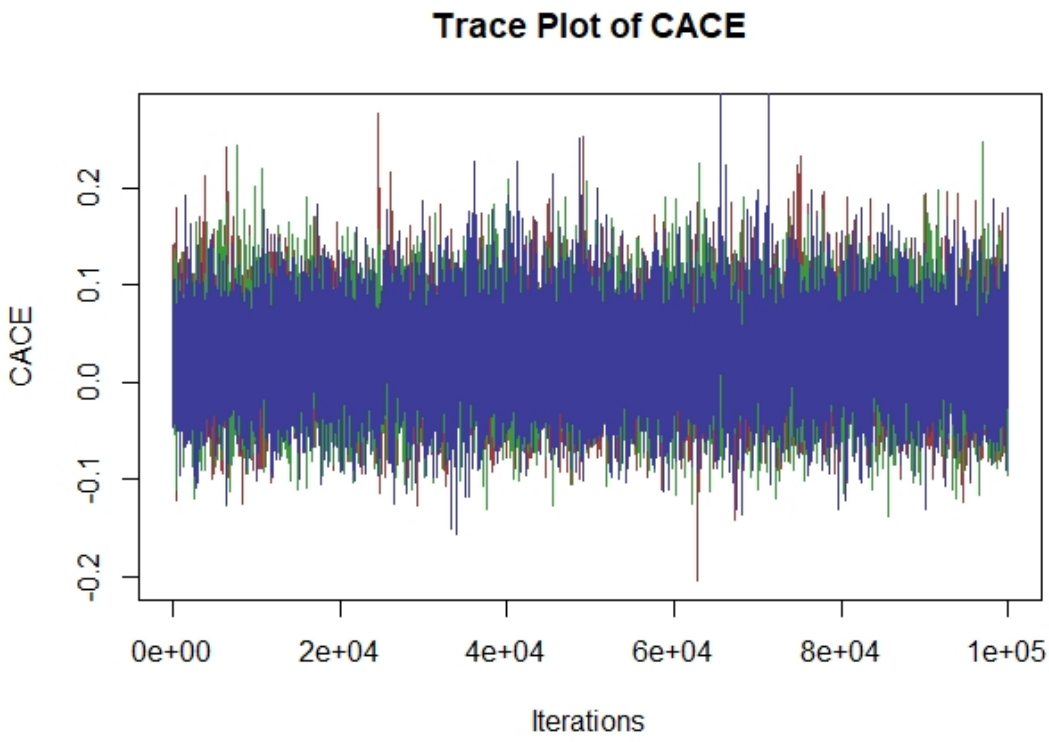

Figure 4.2: Trace plot for θ^{CACE} on the `epidrual_ic` dataset from the model `cace.meta.ic()`.

The trace plot, Figure 4.2, shows the parameter values sampled at each iteration versus the iteration number. Each color represents one chain. Here we used the default `n.chains = 3`. The trace plot shows evidence that the posterior samples of θ^{CACE} are drawn from the stationary distribution.

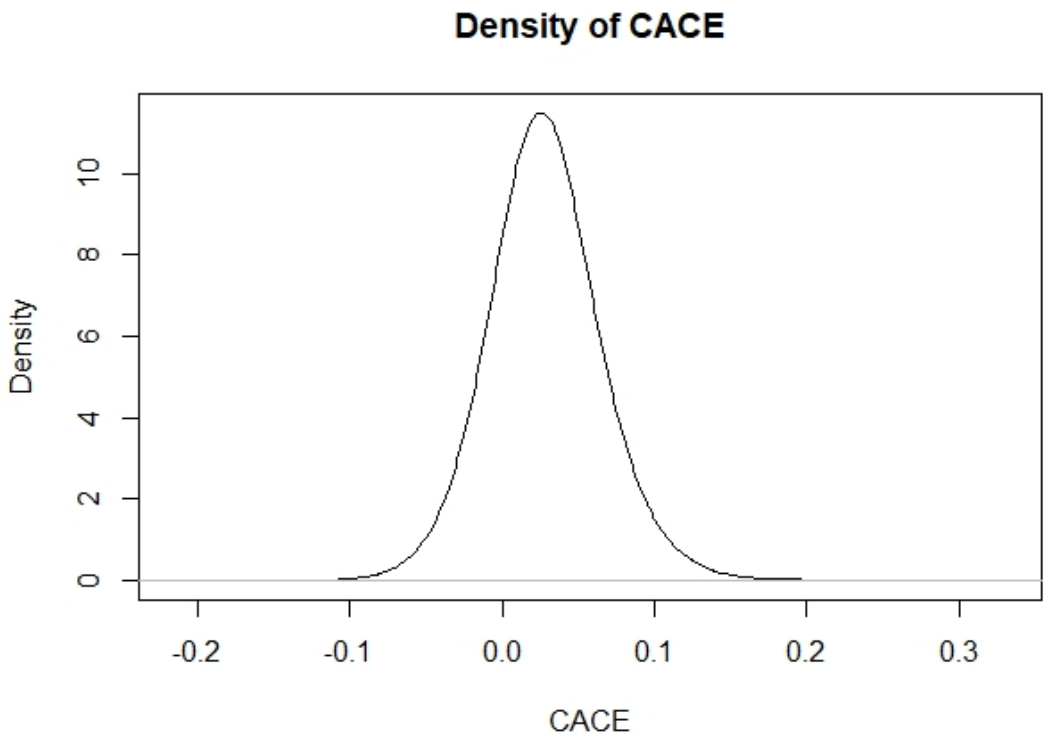

Figure 4.3: The kernel smoothed density for θ^{CACE} from the function `cace.meta.ic()` applied to the epidural analgesia in labor meta-analysis.

The density plot, Figure 4.3, is smoothed using the R function `density()`. It shows that the kernel-smoothed posterior of θ^{CACE} is almost symmetric. The posterior mean is not far from 0, indicating that the complier average causal effect of using epidural analgesia in labor on cesarean section is likely not significant.

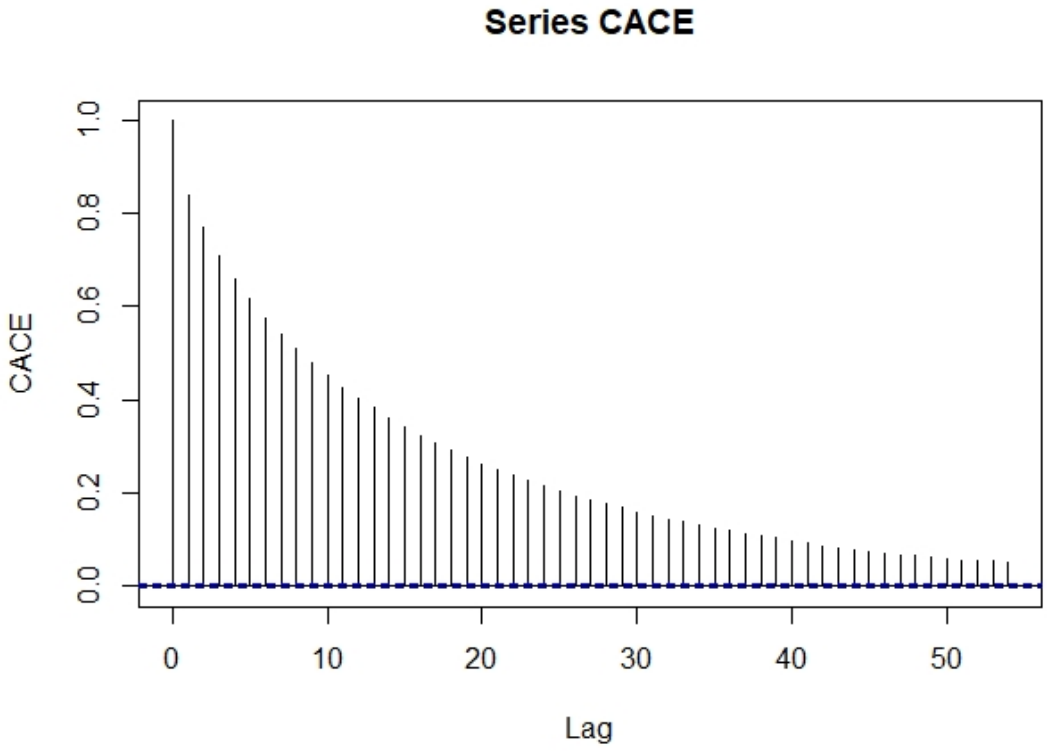

Figure 4.4: Auto-correlation plot of θ^{CACE} from the model `cace.meta.ic()` fit to the `epidrual_ic` dataset.

The autocorrelation plot, Figure 4.4, is a bar plot displaying the auto-correlation for different lags. At lag 0, the value of the chain has perfect auto-correlation with itself. As the lag becomes greater, the values become less correlated. After a lag of about 50, the auto-correlation drops below 0.1. If the plot shows high auto-correlation, users can run the chain longer or can choose a larger `n.thin`, e.g., `n.thin=10` would keep only 1 out of every 10 iterations, so that the thinned out chain is expected to have the auto-correlation dropping quickly.

4.3.5 Plotting the study-specific CACE in a forest plot

A graphical overview of the results can be obtained by creating a forest plot (Lewis and Clarke, 2001). The function `plot.forest()` draws a forest plot for θ^{CACE} estimated from the meta-analysis. Users can call this function for the objects from `cace.meta.c()` or `cace.meta.ic()`. Here is an example using the object `out.meta.ic`:

```
R> plot.forest(data = epidural_ic, obj = out.meta.ic)
```

Note that in addition to the object `out.meta.ic`, users also need to specify the dataset used to compute that object, from which the `plot.forest()` function extracts the study names and publication years for the figure.

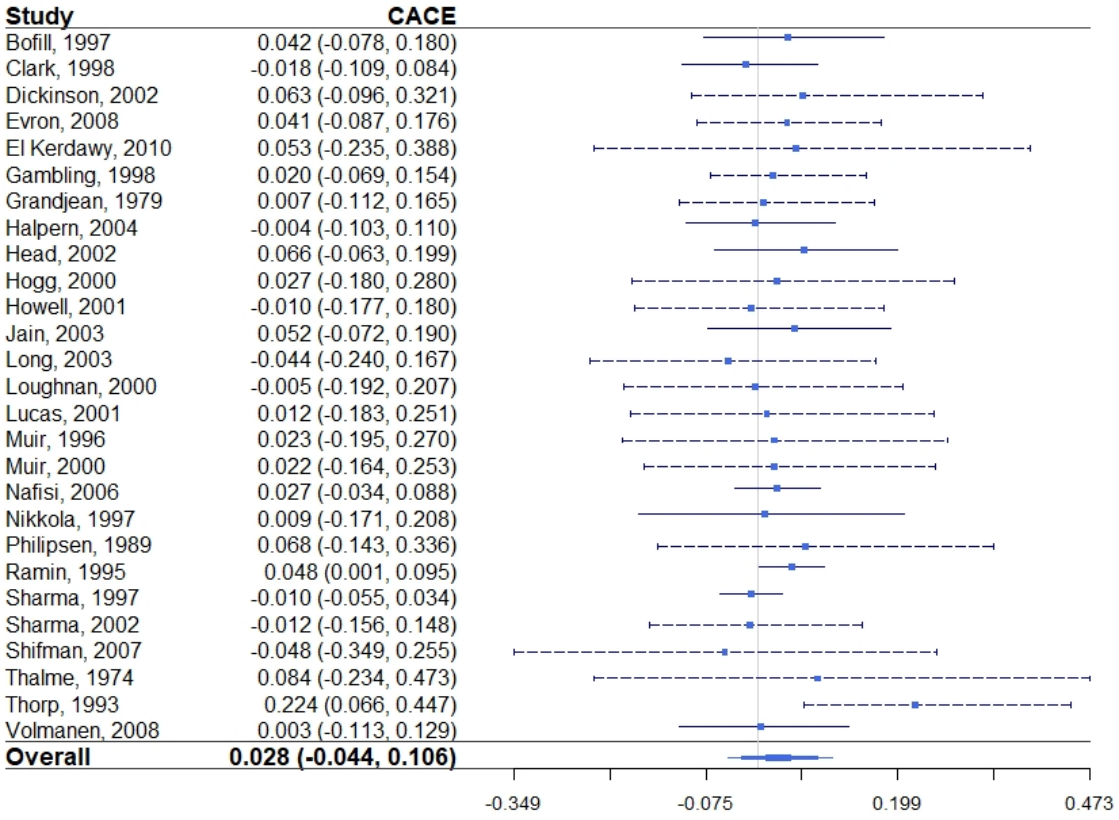

Figure 4.5: Forest plot of study-specific θ^{CACE} from the model `cace.meta.ic()` with full random effects fit to the `epidural_ic` dataset.

Figure 4.5 is a forest plot of θ_i^{CACE} for each study individually, using the Bayesian method with full random effects and default priors. The summary estimate based on the model `cace.meta.ic()` is automatically added to the figure, with the outer edges of the polygon indicating the confidence interval limits. The 95% credible interval of the summary θ^{CACE} covers zero, indicating a non-significant complier average causal effect estimate for using epidural analgesia in labor on the risk of cesarean section for the meta-analysis with 27 trials. For a study with incomplete data on compliance status, a dashed horizontal line in the forest plot is used to represent the posterior 95% credible interval of θ_i^{CACE} from the Bayesian hierarchical model fit. The study-specific θ_i^{CACE}

vary from negative to positive in individual studies, while most of the 95% credible intervals cover zero. As the θ_i^{CACE} for a trial without complete compliance data is not estimable using only data from that single trial, dashed lines tend to have longer credible intervals than those with complete data (solid lines).

If neither random effect δ_{iu} nor δ_{iv} is included in the model (`delta.u = FALSE, delta.v = FALSE` in the calls to `cace.meta.c()` or `cace.meta.ic()`), the study-specific $\theta_i^{\text{CACE}} = u_{i1} - v_{i1}$ are the same across trials, so a forest plot cannot be made. However, the function `cace.study()` can estimate CACE separately for an individual trial as long as it has complete compliance data. In that case, users can choose to generate a forest plot θ_i^{CACE} for each individual study based on separate analyses. The pooled estimate of θ^{CACE} and its 95% credible interval or confidence interval (the diamond in the plot) can be either from the Bayesian hierarchical model (the `cace.meta.c()` function) or from the two-step approach `cace.study()` function with the argument `two.step = TRUE`). The following code is an example of how to create such a plot:

```
R> plot.forest(data = epidural_c, obj = out.study, obj2 = out.meta.c)
```

If `obj` contains the two-step meta-analysis result, the argument `obj2` is optional and is included if the user wants to report the summary CACE estimate based on the Bayesian hierarchical model `cace.meta.ic()`.

4.4 Discussion

This chapter is meant to provide an overview of the **BayesCACE** package for conducting CACE analysis with R. Bayesian hierarchical models estimating the CACE in individual studies and in meta-analysis are introduced to demonstrate the underlying methods of the functions. Practical usage of various functions is illustrated using real meta-analyses datasets `epidural_c` and `epidural_ic`. The package provides several plots for interpretation of model outputs and model diagnosis.

It is important to note that the two-step approach for meta-analysis is included in the package **BayesCACE** because by using the full observed data from a single study i, θ_i^{CACE}

is identifiable, making it possible to pool the estimated posterior means and standard deviations of the θ_i^{CACE} in a meta-analysis. However, the Bayesian hierarchical-model meta-analysis method for estimating the overall CACE, introduced in Sections 4.2.2 and 4.2.3, is preferred for two reasons: the conventional two-step approach requires the whole set of parameters to be estimated for each trial, giving a larger total number of parameters than the random effect model, so the estimate of the CACE can be less efficient; and when study i does not report complete compliance data, it must be excluded from the two-step approach because θ_i^{CACE} is no longer directly estimable by simply using the incomplete data from this individual study, while the function `cace.meta.ic()` can use the incomplete information and thus help to improve efficacy in estimation.

The Gelman and Rubin convergence statistics, time-series standard errors, trace plots, and auto-correlation plots are provided by the package **BayesCACE** to examine whether the MCMC chains are drawn from stationary distributions. However, in practice, any sample is finite, there is no guaranteed way to prove that the sampler has converged (Kass et al., 1998; Cowles and Carlin, 1996). Additional techniques may be required to determine the effective sample size for adequate convergence (Robert and Casella, 2013). For example, the well-developed R package **mcmcse** (Flegal et al., 2017) can be used to assess whether MCMC has been run for enough iterations (sufficient chain lengths). To call the functions in **mcmcse**, users can specify the argument `mcmc.samples = TRUE` in `cace.study()`, `cace.meta.c()`, and `cace.meta.ic()`, so the MCMC posterior samples of monitored parameters are saved in the output objects.

The current version of **BayesCACE** only applies to binary outcomes. However, the Bayesian hierarchical model can be extended to handle ordinal outcomes $o \in \{1, \ldots, O\}$. By selecting weighting scores $\{W_1, W_2, \ldots, W_O\}$ to reflect distances between outcomes categories $\{1, \ldots, O\}$, θ_i^{CACE} is defined as $E(Y_{ij}^1 - Y_{ij}^0 | C_{ij} = 1) = \sum_o (W_o \times u_{io}) - \sum_o (W_o \times v_{io})$ (Zhou et al., 2019). Equally spaced scores $\{1, 2, ..., O\}$, their linear transforms, and midranks are reasonable weight choices (Agresti, 2003). Future work will add CACE meta-analysis functions for ordinal outcomes, and allow users to choose their preferred weights $\{W_1, W_2, \ldots, W_O\}$. Note that ordinal outcomes lead to more complex

correlation structures in the parameters related to response rates, so multivariate prior distributions are necessary to analyze such outcomes. Functions to handle ordinal outcomes and various random effects options are currently under development and may be included in the package at a later point.

Chapter 5

Conclusion

5.1 Summary of major findings

The major contribution of this thesis is that it developed comprehensive Bayesian methods to estimate CACE in meta-analysis of RCTs with binary or ordinal outcomes, accounting for between-study heterogeneity in noncompliance, or both heterogeneous and incomplete noncompliance among studies. To implement this method for binary outcomes, a user-friendly R package **BayesCACE** is provided, where flexible functions for analyzing data from a single RCT and from a meta-analysis of multiple RCTs with either complete or incomplete noncompliance data are developed based on the proposed Bayesian hierarchical models.

In Chapter 2, the proposed Bayesian hierarchical models that account for the inherent heterogeneities between studies and treatment groups in noncompliance were applied to a real meta-analysis of epidural analgesia trials. The case study suggested that including appropriate random effects improved the model fit (according to deviance information criterion) and including different random effects affected the estimates of parameters. Simulation studies were also performed to evaluate the performance of our approach and the impact of misspecification of random effects. The simulations showed that our approach had a good chance of identifying the correct model. When over-fitting occurred, the credible intervals were noticeably too wide and the coverage

probabilities were larger than the nominal level, while under-fitting could produce meaningful decrements in coverage probabilities. To the best of our knowledge, this is the first meta-analysis of RCTs estimating CACE accounting for noncompliance.

The above approach was extended to account for both heterogeneous and incomplete noncompliance among studies in Chapter 3. The method was applied to the same case study of epidural analgesia trials. Using the proposed method, all 27 epidural analgesia trials were included in the CACE meta-analysis. After including the information from studies with incomplete data, the final model we selected had one more random effect than when we applied the model to studies with complete data only, indicating that the added studies may have introduced more heterogeneity and may affect the CACE estimation. We then conducted simulation studies to evaluate the performance of our approach under different missingness mechanisms, and the impact of misspecification of random effects.

Chapter 4 is an application chapter; it introduced the **BayesCACE** package for conducting CACE analysis with R. Besides meta-analysis, Bayesian hierarchical models estimating the CACE in individual studies were also reviewed to demonstrate the underlying methods of the functions. Real meta-analysis datasets used to illustrate practical usage of various functions in **BayesCACE** were `epidural_c` and `epidural_ic`, which corresponded to the case studies analyzed in Chapter 2 and 3. The package also provided various plot functions for interpretation of model outputs and model diagnosis. The upcoming publication of the **BayesCACE** package will allow users to implement the proposed methods and conduct CACE analysis easily and accurately. We believe this newly-developed user-friendly statistical tool will be helpful to popularize CACE meta-analysis methods.

5.2 Future research

Despite the contributions of the Bayesian methods proposed for CACE analysis in meta-analysis, we plan to pursue several future developments. In particular, inspired by the discussions in Chapter 2-4, the following topics on causal inference in meta-analysis need

development.

(1) *Extensions to handle more complicated noncompliance and other types of outcomes.* Throughout the thesis, we only considered all-or-none noncompliance (a dichotomous status) for binary or ordinal outcomes. However, in some trials it is not practical to only discuss all-or-none noncompliance. For example, if the intervention of a RCT is taking a new medicine versus placebo for 6 weeks, patients' compliance status can change at different follow-up time points, so it is more reasonable to consider noncompliance as a time dependent variable. Accounting for such a variable will increase the complexity of defining and estimating causal effects. Recently, extensions have been developed for estimating causal effects in a single study, *e.g.*, causal inference on other types of outcomes such as longitudinal endpoints (Yau and Little, 2001; Pickles and Croudace, 2010) and time-to-event data (Baker and Kramer, 2005; Lok et al., 2004; Yang et al., 2012). In meta-analysis, potential future topics can be chosen from extending the above causal inference research progress from a single study to meta-analysis.

(2) *Relaxing certain key assumptions.* The CACE framework is valid under five essential assumptions: SUTVA, random assignment, the exclusion restriction, $E[T_{ij}^1 - T_{ij}^0] \neq 0$, and monotonicity. These assumptions are plausible in our epidural analgesia case studies, as discussed in Section 2.2.1. However, some assumptions may not be satisfied in other applications. Research on relaxing certain key assumptions has been developed for a single randomized trial, including estimating causal effects in the presence of interference (Hudgens and Halloran, 2008; Liu and Hudgens, 2014) and noncompliance measured with error (Imai and Yamamoto, 2010), *etc.*. Extensions to meta-analysis accounting for these assumptions violations await further development.

(3) *Causal inference in network meta-anlaysis.* Network meta-analysis has recently attracted much attention in evidence-based medicine. It expands the scope of a conventional pairwise meta-analysis to simultaneously compare multiple treatments by synthesizing both direct and indirect information (Lumley, 2002; Zhang

et al., 2014). It can also be used to select the best treatment(s) and to estimate the uncertainty in the treatment ranking and thus facilitates better decision making. Extending the CACE meta-analysis methods to network meta-analysis is also a promising future research topic that awaits further exploration.

(4) *Future work for user-friendly software.* As discussed in Section 4.4, the current version of the **BayesCACE** package provides extensive options for fitting different CACE models, but the functions only apply to binary outcomes. Our proposed Bayesian hierarchical methods can also handle ordinal outcomes, though they bring more complex correlation structures for the parameters and require multivariate prior distributions. We will continue to develop this R package to further expand on what is currently available. Future work would add functions for ordinal outcomes and various random effects options. In the meantime, a frequentist alternative method via the generalized linear mixed model (GLMM) can also be applied to estimate CACE in meta-analysis, and these models are easy to implement in SAS. In addition to the Bayesian approach available from the proposed R package, we plan to release user-friendly SAS macros with a tutorial paper so that researchers who prefer frequentist methods can also implement the CACE analysis easily.

References

Agresti, A. (2003). *Categorical Data Analysis*, volume 482. John Wiley & Sons. 11, 12, 37, 93

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91,** 444–455. 9, 34

Baker, S. G. (2011). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome: an alternative approach. *Biometrics* **67,** 319–323. 10, 35

Baker, S. G. and Kramer, B. S. (2005). Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research* **14,** 349–67. 3, 4, 97

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* **92,** 1171–1176. 2

Bannister-Tyrrell, M., Miladinovic, B., Roberts, C. L., and Ford, J. B. (2015). Adjustment for compliance behavior in trials of epidural analgesia in labor using instrumental variable meta-analysis. *Journal of Clinical Epidemiology* **68,** 525–533. 3, 13, 14, 27, 28, 30, 70, 71

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7,** 434–455. 76, 78

Celeux, G., Forbes, F., Robert, C. P., Titterington, D. M., et al. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1,** 651–673. 13

Chen, H., Geng, Z., and Zhou, X.-H. (2009). Identifiability and estimation of causal effects in randomized trials with noncompliance and completely nonignorable missing data. *Biometrics* **65,** 675–682. 4, 60

Cheng, J. (2009). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics* **65,** 96–103. 2, 10, 35

Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* **59,** 1331–1332. 14, 30

Clopper, C. J. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26,** 404–413. 14, 73

Coggeshall, S. (2017). *noncomplyR: Bayesian Analysis of Randomized Experiments with Non-compliance*. R package version 1.0. 5

Cowles, M. K. and Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91,** 883–904. 93

Egger, M., Davey-Smith, G., and Altman, D. (2008). *Systematic Reviews in Health Care: Meta-analysis in Context*. John Wiley & Sons. 4

Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2017). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN. R package version 1.3-2. 13, 93

Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* **86,** 365–79. 1

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58,** 21–29. 2

Freedman, L. S. (1990). The effect of partial noncompliance on the power of a clinical trial. *Controlled Clinical Trials* **11,** 157–168. 1

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409. 12, 40

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7,** 457–472. 13, 41, 76, 78

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. 13

Glasziou, P. (1992). Meta-analysis adjusting for compliance: the example of screening for breast cancer. *Journal of Clinical Epidemiology* **45,** 1251–1256. 2

Gordon, M. and Lumley, T. (2017). *forestplot: Advanced Forest Plot Using 'grid' Graphics*. R package version 1.7.2. 70

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72,** 320–338. 76

Hedges, L. V. and Olkin, I. (2014). *Statistical Methods for Meta-analysis*. Academic press. 66

Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods* **3,** 486. 66

Higgins, J. and Thompson, S. G. (2002). Quantifying heterogeneity in a metaâĂŘanalysis. *Statistics in Medicine* **21,** 1539–1558. 14

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81,** 945–960. 1

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association* **103,** 832–842. 2, 97

Imai, K. and Yamamoto, T. (2010). Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis. *American Journal of Political Science* **54,** 543–560. 25, 97

Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* pages 305–327. 2, 8, 33

Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine* **30,** 2481–2498. 4, 11

Jiang, Y. and Small, D. (2014). *ivpack: Instrumental Variable Estimation.* R package version 1.2. 5

Jo, B., Ginexi, E. M., and Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prevention Science* **11,** 384–396. 4

Kasim, A., Xiao, Z., Higgings, S., and De Troyer, E. (2017). *eefAnalytics: Analysing Education Trials.* R package version 1.0.6. 5

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain monte carlo in practice: a roundtable discussion. *The American Statistician* **52,** 93–100. 93

Laird, N. M. and Mosteller, F. (1990). Some statistical methods for combining experimental results. *International Journal of Technology Assessment in Health Care* **6,** 5–30. 66

Lewis, S. and Clarke, M. (2001). Forest plots: trying to see the wood and the trees. *BMJ* **322,** 1479–1480. 90

Lian, Q., Hodges, J. S., and Chu, H. (2018). A bayesian hierarchical summary receiver operating characteristic model for network meta-analysis of diagnostic tests. *Journal of the American Statistical Association* pages 1–13. 4

Lin, D. and Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* **97,** 321–332. 42, 67

Little, R. J., Long, Q., and Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* **65,** 640–649. 1

Little, R. J. and Rubin, D. B. (2014). *Statistical Analysis with Missing Data.* John Wiley & Sons. 49

Little, R. J. and Yau, L. H. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using rubin's causal model. *Psychological Methods* **3,** 147. 1, 2

Liu, L. and Hudgens, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association* **109,** 288–301. 2, 25, 97

Lok, J., Gill, R., Van Der Vaart, A., and Robins, J. (2004). Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica* **58,** 271–295. 97

Louis, T. A. and Zeger, S. L. (2009). Effective communication of standard errors and confidence intervals. *Biostatistics* **10,** 1–2. 46, 108

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine* **21,** 2313–2324. 4, 60, 97

Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., and Thomas, A. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis.* Chapman and Hall/CRC. 17, 43, 84

Ma, X., Lian, Q., Chu, H., Ibrahim, J. G., and Chen, Y. (2018). A bayesian hierarchical model for network meta-analysis of multiple diagnostic tests. *Biostatistics* **19,** 87–102. 4

Natarajan, R. and McCulloch, C. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* **7,** 267–277. 16

Newcombe, R. (1988). Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine* **7,** 1179–1186. 3

Nikkola, E. M., Ekblad, U. U., Kero, P. O., Alihanka, J. J., and Salonen, M. A. (1997). Intravenous fentanyl PCA during labour. *Canadian Journal of Anaesthesia* **44,** 1248–1255. 14

O'Malley, A. J. and Normand, S.-L. T. (2005). Likelihood methods for treatment noncompliance and subsequent nonresponse in randomized trials. *Biometrics* **61,** 325–334. 4

Peng, Y., Little, R. J., and Raghunathan, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics* **60,** 598–607. 4, 60

Pickles, A. and Croudace, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research* **19,** 271–289. 97

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna, Austria. 31

Plummer, M. (2018). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-8. 70, 75, 86

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News* **6,** 7–11. 70

Riley, R. D., Jackson, D., Salanti, G., Burke, D. L., Price, M., Kirkham, J., and White, I. R. (2017). Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* **358,** j3932. 4

Robert, C. and Casella, G. (2013). *Monte Carlo Statistical Methods*. Springer Science & Business Media. 93

Roy, J., Hogan, J. W., and Marcus, B. H. (2007). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* **9,** 277–289. 25

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association* **75,** 591–593. 9, 34

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 583–639. 13, 41, 76

Stuart, E. A., Perry, D. F., Le, H.-N., and Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science* **9,** 288–298. 4

Ueberhuber, C. W. (1997). *Numerical Computation 1: Methods, Software, and Analysis*, volume 16. Springer Science & Business Media. 46, 108

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30,** 261–293. 24

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36,** 1–48. 66, 76

White, H. (1982). Instrumental variables regression with independent observations. *Econometrica: Journal of the Econometric Society* pages 483–499. 5

Yang, Z., Brentnall, A. R., Cuzick, J., and Sasieni, P. (2012). Estimating the complier average causal effect for exponential survival in the presence of mid-trial switching. *International Journal of Statistics in Medical Research* **1,** 84–90. 60, 97

Yau, L. H. and Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* **96,** 1232–1244. 1, 60, 97

Zeger, S., Liang, K., and Albert, P. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44,** 1049–1060. 16, 41, 67

Zhang, J., Carlin, B. P., Neaton, J. D., Soon, G. G., Nie, L., Kane, R., Virnig, B. A., and Chu, H. (2014). Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clinical Trials* **11,** 246–262. 4, 60, 97

Zhang, J., Chu, H., Hong, H., Virnig, B. A., and Carlin, B. P. (2017). Bayesian hierarchical models for network meta-analysis incorporating nonignorable missingness. *Statistical Methods in Medical Research* **26,** 2227–2243. 50

Zhou, J., Hodges, J. S., Suri, M. F. K., and Chu, H. (2019). A Bayesian hierarchical model estimating CACE in meta-analysis of randomized clinical trials with noncompliance. *Biometrics* . 5, 6, 63, 64, 66, 71, 93

Appendix A

Supplementary Materials for Chapter 2

A.1 Data display for the case study

Table A1 shows the frequencies of cesarean deliveries, based on the allocated intervention and the actual received analgesia, for the 10 trials analyzed in Section 3. In the table, n_{irto} refers to observed count for outcome o in the i^{th} study group $\{j : R_{ij} = r,\ T_{ij} = t\}$, $o \in \{0,\ 1\}$.

Table A1: Data in 10 randomized controlled trials of epidural analgesia in labor

Study (Author, Year)	Allocated control				Allocated epidural			
	Received control		Received epidural		Received control		Received epidural	
	Cesarean – n_{i000}	+ n_{i001}	Cesarean – n_{i010}	+ n_{i011}	Cesarean – n_{i100}	+ n_{i101}	Cesarean – n_{i110}	+ n_{i111}
Bofill, 1997	37	2	11	1	2	0	42	5
Clark, 1998	72	6	68	16	7	2	134	13
Halpern, 2004	62	5	44	7	0	0	112	12
Head, 2002	51	7	2	0	3	0	43	10
Jain, 2003	72	11	0	0	0	2	36	7
Nafisi, 2006	179	19	0	0	0	0	173	24
Nirkola, 1997	6	0	4	0	0	0	10	0
Ramin, 1995	546	17	95	8	230	0	393	39
Sharma, 1997	336	16	5	0	114	1	231	12
Volmanen, 2008	23	1	3	0	1	0	23	1

A.2 Parameter estimates of the case study

Table A2 presents parameter estimates from the fixed effects model (Model I) and the final model (Model IVa) in Section 3.3. We used the triple of percentiles, ${}_{2.5}50_{97.5}$, to display a parameter posterior median with its 95% equal tail credible interval, as suggested by Louis and Zeger (2009). Monte Carlo integration (Ueberhuber, 1997) was used to estimate the overall probabilities of being a principal stratum, π_a, π_c, and π_n when δ_{in} and δ_{ia} were both present (Model IVa), and the overall never-taker response rate, $s_1 = E(s_{i1})$ of Model IVa was estimated using the approximation $E(s_{i1}) \approx \text{logit}^{-1}(\frac{\alpha_s}{\sqrt{1+C^2\sigma_s^2}})$, $C = \frac{16\sqrt{3}}{15\pi}$. Estimates for other overall response rates, b_1, u_1, v_1, and θ^{CACE} were directly from the fixed parameters because no random effect on those rates was included in either Model I or Model IVa. For example, b_{i1} were assumed to be the same across studies according to the models so we estimated $b_1 = E(b_{i1}) = \text{logit}^{-1}(\alpha_b)$. Based on the final Model IVa, the posterior estimate of θ^{CACE} was ${}_{-0.003}0.022_{0.048}$, which covered zero and indicated a nonsignificant complier average causal effect of epidural analgesia in labor on cesarean section. The random effects of n_i and a_i had standard deviations of 1.619 and 1.912, while the random effect of s_1 had a standard deviation of

2.004 on the logit scale. After adding random effects δ_{ia}, δ_{in} and δ_{is}, the estimates of the mean parameters of n_i and a_i from Model IVa were much different from the estimates from Model I.

Table A2: Summary of parameter estimates using Bayesian hierarchical models for 10 RCTs of epidural analgesia in labor

Parameter	Model I (None)	Model IVa ($\delta_{ia}, \delta_{in}, \delta_{is}$)
θ^{CACE}	$_{-0.011}0.016_{0.044}$	$_{-0.003}0.022_{0.048}$
Overall never-taker probability π_n	$_{0.197}0.216_{0.236}$	$_{0.083}0.096_{0.116}$
Overall always-taker probability π_a	$_{0.136}0.153_{0.170}$	$_{0.155}0.186_{0.223}$
Overall complier probability π_c	$_{0.604}0.631_{0.656}$	$_{0.677}0.717_{0.752}$
Overall never-taker response s_1	$_{0.010}0.021_{0.039}$	$_{0.050}0.172_{0.455}$
Always-taker response b_1	$_{0.087}0.123_{0.166}$	$_{0.082}0.113_{0.148}$
Treated complier response u_1	$_{0.065}0.086_{0.108}$	$_{0.072}0.091_{0.111}$
Control complier response v_1	$_{0.053}0.069_{0.087}$	$_{0.053}0.069_{0.086}$
Mean parameter of n_i	$_{-1.197}-1.070_{-0.945}$	$_{-4.154}-2.828_{-1.716}$
Mean parameter of a_i	$_{-1.562}-1.419_{-1.280}$	$_{-3.452}-2.168_{-0.893}$
Standard deviation of n_i	–	$_{1.024}1.619_{2.888}$
Standard deviation of a_i	–	$_{1.235}1.912_{3.303}$
Standard deviation of logit(s_{i1})	–	$_{1.061}2.004_{4.354}$

The triple notation of ${}_LP_U$ denotes the posterior median P with 95% equal tailed credible limits (L, U) using Bayesian hierarchical models.

A.3 Sample R code

Sample R JAGS code for Model IVa is shown below. More modeling and simulation codes are available at our GitHub repository: https://github.com/JinchengZ/CACE_meta.

```
model{
for (i in 1:n) {
```

```
prob[i, 1] <- (pi_n[i]*(1-s1[i]) + pi_c[i]*(1-v1[i]))
prob[i, 2] <- (pi_n[i]*s1[i] + pi_c[i]*v1[i])
prob[i, 3] <- (pi_a[i]*(1-b1[i]))
prob[i, 4] <- (pi_a[i]*b1[i])
prob[i, 5] <- (pi_n[i]*(1-s1[i]))
prob[i, 6] <- (pi_n[i]*s1[i])
prob[i, 7] <- (pi_c[i]*(1-u1[i])+pi_a[i]*(1-b1[i]))
prob[i, 8] <- (pi_c[i]*u1[i]+pi_a[i]*b1[i])

R[i, 1:4] ~ dmulti(prob[i, 1:4], N0[i])
R[i, 5:8] ~ dmulti(prob[i, 5:8], N1[i])

probit(u1[i]) <- alpha_u
probit(v1[i]) <- alpha_v

n[i] <- alpha_n + delta_n[i]
delta_n[i] ~ dnorm(0, tau_n)
a[i] <- alpha_a + delta_a[i]
delta_a[i] ~ dnorm(0, tau_a)
pi_n[i] <- exp(n[i])/(1+exp(n[i])+exp(a[i]))
pi_a[i] <- exp(a[i])/(1+exp(n[i])+exp(a[i]))
pi_c[i] <- 1-pi_a[i]-pi_n[i]

logit(s1[i]) <- alpha_s1 + delta_s[i]
delta_s[i] ~ dnorm(0, tau_s)
logit(b1[i]) <- alpha_b1
}

CACE <- phi(alpha_u)-phi(alpha_v)
```

```
# priors
alpha_n ~  dnorm(0, 0.16)
tau_n ~ dgamma(2, 2)
sigma_n <- 1/sqrt(tau_n)
alpha_a ~ dnorm(0, 0.16)
tau_a ~ dgamma(2, 2)
sigma_a <- 1/sqrt(tau_a)

alpha_s1 ~  dnorm(0, 0.25)
tau_s ~ dgamma(2, 2)
sigma_s <- 1/sqrt(tau_s)
alpha_b1 ~  dnorm(0, 0.25)
alpha_u ~  dnorm(0, 0.25)
alpha_v ~  dnorm(0, 0.25)
}
```

A.4 Sensitivity analysis on priors in the case study

In a sensitivity analysis of the prior specification, we repeated our analysis of the final model (Model IVa) using two different sets of priors for the variance parameters of random effects, which were more diffuse than the ones presented earlier. Specifically, we considered $Gamma(1,1)$ and $Gamma(0.5,0.5)$ as priors for σ_n^{-2}, σ_a^{-2}, and σ_s^{-2}, which correspond to a 95% interval for the variance parameters of $(0.27, 39.50)$ and $(0.2, 1018.3)$, respectively. The posterior estimate of θ^{CACE} was ${}_{-0.0035}0.0223_{0.0483}$ when using $Gamma(1,1)$, and ${}_{-0.0032}0.0223_{0.0483}$ under $Gamma(0.5,0.5)$. Comparing the result ${}_{-0.0034}0.0222_{0.0481}$ from Model IVa, the posterior estimate of θ^{CACE} did not change to the third digit. Therefore, for the priors considered, the results were consistent.

A.5 Additional simulation results

In addition to the four data generating settings presented in Table 4, we added two additional settings and displayed the mean, bias, 95% credible interval length, and coverage probability for θ^{CACE} under each candidate model. In both scenarios the data were generated from a full model as introduced in Section 2.4, with all of the 6 possible random variables included. In the first setting we used the probit link function for u_{i1} and v_{i1}, while in the second scenario a logit link was used. Sample sizes, allocation ratio, and true values were set the same as described in Section 4. The variances of δ_{in}, δ_{ia}, δ_{is}, δ_{ib}, δ_{iu} and δ_{iv} were set to 0.5^2, and the true correlation between δ_{in} and δ_{ia} was 0.5. Besides the 8 candidate models with up to 3 random effects, we also fitted the generated data with the full model (6 random effects, probit link for u_{i1} and v_{i1}).

As θ^{CACE} is only related with u_{i1} (the response rate of a treated complier) and v_{i1} (the response rate of complier in control), failure to include the random effect for δ_{iu} and δ_{iv} reduces the coverage probability for θ^{CACE} substantially. In the meantime, including random effect on δ_{iu} and δ_{iv} tend to give longer 95% credible intervals. In the second setting we simulated heterogeneous θ_i^{CACE} using the logit link for u_{i1} and v_{i1}, but fitted all candidate models with probit link, in order to check if our proposed model still performed well. Results showed that the estimated θ^{CACE} regarding the bias, 95% credible interval length, and coverage probability were pretty much the same as those under the correct link function, which were as expected because the normal and logistic cumulative distribution functions were very close.

Table A3: Additional performance of estimates and credible intervals for θ^{CACE} for each model, based on 2000 simulated datasets

True Random Effects Model		Selected Random Effects Model								
		None	δ_{in}	δ_{iu}	δ_{iv}	δ_{in}, δ_{iu}	δ_{in}, δ_{iv}	δ_{iu}, δ_{iv}	δ_{in}, δ_{iu}, δ_{iv}	All
All probit	Mean	-0.347	-0.350	-0.338	-0.338	-0.341	-0.337	-0.329	-0.328	-0.330
	Bias	-0.002	-0.005	0.007	0.007	0.004	0.008	0.016	0.017	0.015
	95% CIL	0.102	0.102	0.212	0.214	0.212	0.215	0.284	0.285	0.278
	95% CICp	0.647	0.606	0.954	0.961	0.933	0.958	0.989	0.989	0.988
All logit	Mean	-0.233	-0.234	-0.226	-0.227	-0.227	-0.219	-0.220	-0.219	-0.219
	Bias	0.002	0.002	0.009	0.009	0.008	0.016	0.015	0.016	0.016
	95% CIL	0.103	0.102	0.198	0.201	0.197	0.264	0.263	0.264	0.260
	95% CICp	0.764	0.737	0.982	0.983	0.976	0.997	0.997	0.997	0.997

The bold-face cells represent the correctly chosen model.

95% CIL: 95% equal-tail credible interval length.

95% CICp: 95% credible interval coverage probability.